

Upheaval for Soviet academy

The coming of perestroika seems to have caught the Soviet Academy of Sciences at a loss for knowing what to do. Here is how it might tackle some of the questions arising later in the month.

THE Soviet Academy of Sciences will find itself having to cross another Rubicon at its general meeting later this month. The chief business is to decide how to apply to itself, the most important part of the Soviet science establishment, the decisions reached at the Special Conference of the Soviet Communist Party in June. How the academy decides will begin to answer the absorbing question whether the much-needed reform of the research enterprise will be in the van, or in the rearguard, of *perestroika*.

Some things will be simple, of course. It would plainly be out of tune with the decisions of the conference on public office-holding that Academician A. D. Mirzabekov, the able director of the Moscow Institute of Molecular Biology, should keep that job while also serving as the head of the department of the academy (now one out of eighteen) responsible for all research in the field, the programme (and finances) of his own included. Almost certainly, he will decide to remain head of the department, finding somebody else to replace him at the institute. The academy will no doubt also tackle conscientiously the contentious issues of whether septuagenarians and their seniors should remain in important administrative posts, and whether the general principle (as from June) that nobody should serve more than two terms in the same appointment should apply retrospectively to those now in office: the temptation to apply the new rules only to those not yet appointed should be resisted, even though there may be a shortage of younger people qualified to fill the posts thus vacated.

That is but one of the consequences of the simple truth that the academy, with fewer than 300 full members, is both relatively too small, by the scale of the Soviet research establishment, and numerically too small to carry out the work expected of it. But the appointment of more academicians would not resolve the obvious difficulties. Even a ten-fold increase of numbers would still leave the academy awkwardly exclusive, hardly able to claim fairly to reflect the opinions and interests of academic science. And the burdens of duty would remain intolerable. For what the Soviet academy is attempting, sometimes at the behest of the government and sometimes from choice, is not merely a huge task but one that engenders conflicts of interest.

Squeamishness

Mirzabekov, for example, appears to have been quick to appreciate the conflict of interest in becoming the one who decides on the allocation of resources in a field of research while remaining, as an institute director, one of the potential beneficiaries. Many of his predecessors have not been nearly as squeamish. Academician Yuri Ovchinnikov, who died last year, is a memorable but not isolated illustration of how academicians have sometimes allowed their belief in own far-sightedness to still such self-doubts as may have arisen about the wisdom of their generous support of their own interests.

Ovchinnikov, ten years ago an early proselytizer on behalf of biotechnology whose energy led to the building of the production plants for chemicals labelled by radioisotopes and the restriction enzymes on which molecular biology depends, in his

role as grant-maker also endowed himself, an institute director, with physical facilities which are extravagant even by Western standards. The decision may have been wise, but the least that can be said is that it is not transparently so. The account on page 107 of Ovchinnikov's intervention in a project to develop a blood substitute, with tragic consequences (for another), further illustrates how, in research which is necessarily competitive, too great a concentration of power over others in a single person's hands can lead to trouble. The assumption that there is a strong and positive correlation between an academician's administrative power and his current sagacity as a scientist makes matters worse.

To remark on this danger is not to suggest that all powerful academicians in the Soviet Union fall into the same trap. It is simply that the possibility that they may do so shows that the framework in which they operate should be changed. A rule that people should not hold two jobs at the same time would be a start on reform, but only a start.

Difficulty

The academy's essential difficulty is more general, arising from the disparity of its functions — functioning as an academy, giving advice to the government, conducting research through its own institutes, publishing the results — and then shaping the pattern of its research so as to square the circle between researchers' interests and the changing exhortations of the Soviet government. The tasks are complicated by the spread of the academy's intellectual interests (from ethnography to algebra) and the Soviet practices requiring it to provide most of its researchers not merely with equipment for their research but with housing in which they and their families can live.

During the past year, several sensible steps have been taken to simplify the management of these herculean tasks, not the least significant of which is the decentralization of much administrative decision-making to the level of the research institute. It will be especially welcomed in the West that institute directors are now ordinarily the arbiters of whether members of their research staffs should be able to accept invitations to meetings outside the Soviet Union. Foreign travel has even grown to the point at which the academy is hard pressed to pay *Aeroflot* for the airline tickets of all those who wish to travel. But this and other administrative reforms are merely concessions to good management practice which do not reach the heart of the academy's dilemma: on which of its several responsibilities to concentrate?

The best strategy is to resolve that there should ultimately be a separation of functions from the Soviet academy, to decide now what functions should eventually be spun off and to acknowledge that such radical changes cannot be brought about overnight or even by some date specified in advance. The two functions which it is essential to keep intact are those of an academy designed to foster and, when necessary, to defend scholarship and an academy that gives advice to the government. The direct management of research should ultimately be allowed to pass to other organizations. So should the academy's present responsibility for the support of research in general. But



what other organizations? And what mechanisms for supporting research in general should there be? With a little deftness now, the academy could create within its own system a pattern of research management that would serve the interests of Soviet science for decades to come.

The first need is to develop a flexible mechanism for identifying good research projects and then for backing them selectively. What there is at present is merely a cake-cutting mechanism by which institutes are allocated resources within the departments of the academy to which they belong, and which are then carved up internally by a process in which directors often have more influence than they can exercise wisely. But sooner or later there will have to be a Soviet version of what is known in the West as peer-review. Why not sooner rather than later? And why not, because the system is unknown in the Soviet Union, involve researchers from overseas? Some might consider too much pride would be sacrificed in that process, but should remember that the Max-Planck Gesellschaft is but one of several research organizations elsewhere which is unabashed by including foreign scientists in the management of its research. A year ago, the Soviet academy dipped its toe in the water with a scheme for allocating special funds to projects in high-temperature superconductivity, but the more urgent need is to make a peer-review system work in less exceptional hard-core research.

These are among some of the several radical proposals put forward by Dr M. Frank-Kamenetskii in the symposium volume *Ynogo ne dano* ("One way") prepared for distribution before the conference in June, but still difficult to find in Moscow bookshops. But he would go too fast and, in a sense, too furiously, suggesting that only researchers able to win research grants and to have them renewed should keep their jobs, but that would be too rigorous a solution for the problems of over-staffing and immobility with which the research institutes are beset, and which will eventually have to be solved by a transformation of the Soviet economy, and would in any case cause great dissension while there is no basis for confidence in the fairness of a competitive grants system. The better course would be to begin in a restricted but important field of research, waiting five years or more before generalizing the recipe. There is in any case a bigger prize to play for — the possibility that a well-designed system of competitive research grants might break down the barriers that at present isolate academy institutes from the universities on the one hand and the institutes of other academies (health and agriculture in particular) on the other. If the Soviet Academy of Sciences could make competitive grants work well within its own framework, it could surely persuade the other academies and the education ministries to throw their own interests into a common pool.

The other important question the Soviet academy must decide is that of how its research institutes should eventually be disposed. There can be no question of devising now a single recipe that can be carried through even over a period of several years. But it would help if the scientific programmes of the research institutes were overseen less hierarchically than now, by groups of interested and knowledgeable people reporting independently of the institute director to the academy. Such a system might enliven many of the institutes now half-asleep and would certainly avoid the danger that an eccentric director might dragoon his colleagues into wayward pursuits. Independent invigilation of this kind would also enable the academy to decide which of its institutes should eventually be spun off in different directions — towards universities, towards industrial enterprises as they will be in the still unknown Soviet future, or towards free-standing research organizations.

For the Soviet academy as it stands, none of this requires immediate upheaval, which would in any case be inappropriate at a time of great uncertainty. Rather, the need is for a resolve to explore new ways of managing research and a statement of objectives, however distant. Competitive research grants and eventually devolved research institutes are only some of the

stratagems that might be attempted. For the Soviet academy, and thus for Soviet research as a whole, doing nothing is to run the biggest risk. And if the end result is an academy that has shed managerial responsibilities and whose members individually lay less claim to high prestige, the academy as a whole may be able to claim the credit for having brought to life a system of research whose potential is at present only partly realized. □

European science

An enterprise to champion European science and the European spirit should not expect too much.

As predicted, this year's meeting of the British Association for the Advancement of Science at Oxford (see p. 104) seems to be turning out well, suggesting that Britain may, against expectations, have seen a revival of public interest in science precisely when the interest of the British government seems to have reached a low point and when the interests of young people are also declining, at least as measured by the figures for successful applicants for places on university science-based courses. Sir Walter Bodmer, this year's president of the association, is right to have emphasized in his address this week the economic importance of a sufficiently large pool of skilled people, to have deplored public indifference towards science and to have asked how government policy might be changed. His answer is that matters would improve if there were a wider appreciation of the interest and excitement of science. May there be another?

The issue is important because of another enterprise in which Bodmer also has a hand, a scheme to launch at Maastricht in The Netherlands on Saturday a new organization called Foundation Scientific Europe, with long-term objectives very similar to those of the British Association. The idea is to make propaganda to suggest that Europe is bursting with innovation, intellectual and technical. There will be a book extolling past and present achievements "in at least five languages" and there are plans to broadcast television programmes, at least to investigate the feasibility of a popular science magazine and eventually to run a peripatetic European science festival of which a meeting organized along the lines of the British Association would be a part. The project has been pushed along by three distinguished science writers (Nigel Calder, Bernard Dixon and Theo Martens), who evidently hope to use the techniques of communication in two good causes at once — European identity and the public understanding of science. (The immediate need seems to be for money, but the chances should be good, with people such as Bodmer among the backers.)

To judge from the brochures, the new foundation has a fair chance of survival, perhaps even of success. The techniques promise to be very different from those deployed at Oxford this week, but will be none the worse for that. It would not be surprising if the gatherings of people that stirred excitement in early Victorian Britain are inappropriate a century and a half later. It is also entirely within the bounds of possibility that a careful display and explanation of European science will as effectively galvanize European spirits as some more obvious intellectual pursuits — European history, for example. Europe, after all, was where what is now rightly or wrongly called modern science began. But the sponsors should not be disappointed if even the new enterprise does not put everything to rights in the management of science and technology.

For the plain truth is that the fostering of a proper regard for science on the scale modern industry expects inescapably requires attention to the infrastructure on a scale that the voluntary efforts of private citizens cannot sustain. The need begins in elementary schools and continues by means of professional training throughout people's working lives. The research community can help to make their interests accessible and interesting, but governments over-preoccupied with the ratio of public spending to GNP can be decisive. □

International cooperation on superconductors

- First Japanese symposium held
- Nationalistic competition inevitable

Tokyo

JAPANESE superconductor researchers are hoping for international cooperation and the US administration sees the inevitability of nationalistic competition. That was the message from a panel discussion on "international cooperation" by speakers from the United States, Japan, Europe and India at a symposium on superconductivity held in Japan last week.

The symposium was the first arranged by Japan's International Superconductivity and Technology Center (ISTEC), an information centre and research consortium set up earlier this year with funding from the private sector and the Ministry of International Trade and Industry (MITI).

Professor Shoji Tanaka, vice-president of ISTEC, compared the centre's goals with those of other national MITI projects: the very large scale integration (VLSI) project (1976–80); the optoelectronics and optical communications project (1979–85); and the new functional devices project (1981–91).

He pointed out that each new project is longer than the one that went before and orientated towards more distant commercial goals. He expects the present superconductor project to last for about ten years on a budget of around \$150 million, and believes that commercial applications are distant enough to allow international collaboration on basic research.

Setting up collaborative research projects in Japan is not easy. M. Uenohara of NEC Corporation explained that Japanese electronics companies "compete furiously" in the market; at the start of the VLSI project, researchers from the different companies in the project were unwilling to communicate. After a year of negotiations it was agreed to limit research to "pre-competitive technology". From then on the researchers worked together like "family members", according to Uenohara. He thinks a similar style of cooperation on superconductors is possible at the international level.

Not everyone would agree with the value of such international cooperation. The VLSI project has been much studied because of a common perception that it had propelled 'Japan, Inc.' to international dominance of the semiconductor memory market. Many experts feel that the strength of that project (and those that came after) was not in cooperative basic research, but in creating a consensus in giant electronic companies that the semi-

conductor race must be won. So the cooperative project stimulated competition. Foreign companies thus worry that joining in a 'cooperative' project in Japan might simply provide basic research results to fuel the applications race.

Paul McLaughlin, director of the Council on Superconductivity for American Competitiveness (CSAC), said "nationalism" will be the driving force behind advance in superconductor research. But he did acknowledge the need for international exchange of information and says that CSAC's information services are open to everyone.

CSAC is proposing to form a research consortium, the Superchip Corporation, funded by private industry and possibly backed by US federal government bonds. The council is lobbying for legislation to exempt the corporation from US monopoly laws, and Superchip may also get special treatment over patents, McLaughlin says. But unlike Japan's superconductivity centre which is open to foreign companies, CSAC and Superchip are for US companies only.

McLaughlin says the United States will spend \$205 million on superconductor research in the fiscal year beginning 1 October. Much will be spent on defence-related research in government laboratories, which have a poor reputation for transferring technology to the commercial sector. McLaughlin says CSAC's purpose is to ensure that the United States is competitive by organizing communication between government labs and industry.

India, on the other hand, has been hard pushed to scrape up \$10 million for superconductor research this year, according to C. N. R. Rao of the Indian Institute of Science; developing countries "can never be equals" to the advanced nations. And he appealed for international cooperation.

The governments of the United Kingdom, France and West Germany are each spending a few tens of millions of dollars on superconductor research and new research centres have been set up in collaboration with industry, according to panel speakers. Sir Martin Wood, chairman of the UK National Committee for Superconductivity, says a standing committee has been formed by six member nations of the European Community. Wood says he thinks superconductors could provide an important example of European cooperation and "will if I get my way".

David Swinbanks

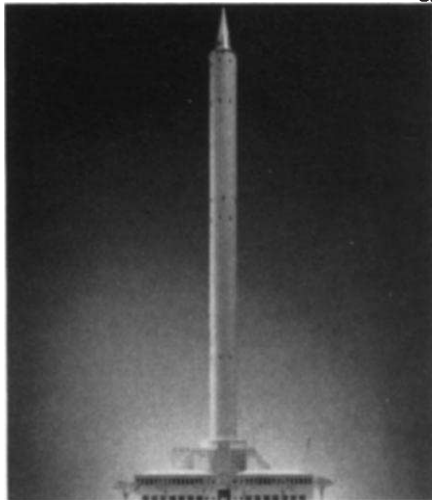
Neural networks

EUROPE'S research project into industrial uses of neural networks begins in November and will cost £3.5 million, half coming from the European information-technology programme ESPRIT and half from industrial collaborators. The project, called ANNIE (application of neural networks for industry in Europe), involves seven laboratories in Britain, France, Germany and Greece, led by the Harwell Laboratory of the UK Atomic Energy Authority. The aim is to compare neural network technology with conventional problem-solving methods.

C. McG.

Microgravity tower

WEST Germany will build a 146-m tower in Bremen to allow experiments to be performed under microgravity conditions. Experiments in combustion, fluid mechanics, materials science and biotechnology



Bremen's planned "drop tower".

will be carried out from late 1989. A capsule will fall 110 m down an evacuated steel tube at the centre of the tower, providing weightless conditions for up to 4.74 seconds. Researchers will also be able to launch capsules from the base of the tower, doubling the period of weightlessness. The Research Ministry will provide DM7 million for the tower, which is being built by the University of Bremen. The city of Bremen, the West German Education Ministry and local industry will bear the rest of the DM18.2 million total costs.

S.D.

Money for Frontiers

JAPAN'S Science and Technology Agency and the Ministry of International Trade and Industry have applied for 2,600 million yen (about \$20 million) for the Human Frontiers Science Program in budget requests submitted to the Ministry of Finance. The funds will be used to establish a foundation in Europe next spring, says Tateo Arimoto of the agency, and the first international grants for research on the brain and molecular recognition and response should be available in 1989.

D.S.

Association flies the science flag

London

THE flag is flying high for science in Britain this week as the 150th meeting of the British Association for the Advancement of Science, known as the BA, gets under way in Oxford. But the theme is not solely British: in his opening speech at the meeting on Monday, the president of the association, Sir Walter Bodmer, called for collaboration on an international scale in order to meet what he sees as one of the most exciting challenges to science now, that is, to map the human genome and create a 'Handbook of Man'.

"There is still a long way to go to decipher the whole of the handbook" he said, "but a coordinated international effort, probably less than that which was required to put a man on the Moon, could perhaps achieve an end result by the year 2000 or soon after. This surely is a challenge we cannot forego. It has immense cultural value, is likely to provide the basis for the prevention or treatment of most major human chronic diseases, and will stimulate technological developments that are certain to be of . . . economic benefit."

The rest of the week, as usual, is packed with enough lectures, exhibitions and tours to exhaust even the fittest of enthusiasts and to frustrate anyone with only a day to spare for a visit. Deciding what to choose from the programme is a task which alone could take up the half a day. "Dangerous", "Puzzles, fudges and convictions", "Magnets, martians and mathematical microworlds", are just some of the lecture titles designed to intrigue visitors.

One of the difficulties in holding such a large number of events in one week is that it takes thousands of visitors to ensure a healthy attendance at each one. And the organizers will be hoping for an improvement on the disappointing attendance at last year's meeting in Belfast.

One lecture that is sure to draw the crowds is being given by Professor Stephen Jay Gould on the myths surrounding the historic clash between orthodox Christianity and Darwinism in the forms of the Bishop of Oxford, Samuel Wilberforce and Thomas Huxley, a debate which took place at the 1860 BA meeting in Oxford. Giving a hint at the content of Gould's talk, a spokesman said Gould would argue that "almost everything that everyone thinks they know about the debate is wrong".

Another event which may attract the crowds with its intriguing title is "Science in the graveyard". Researchers might even be misled by the name into thinking this is the debate on the future of science in Britain. But instead they will find themselves on a geological visit to an urban graveyard.

Christine McGourty

Japan plans to build world's biggest fusion device

Tokyo

JAPAN'S Ministry of Education, Science and Culture last week applied for funds to develop the world's largest helical fusion device, which will be built by a new national institute that will open next April.

Japan's largest fusion facility, the giant JT-60 tokamak, is run by the Science and Technology Agency, but the education ministry also operates several smaller fusion devices in universities around the country, including tokamaks, heliotrons, tandem-mirror and laser devices.

In 1986, the ministry's Science Council decided that a new institute, equipped with a large helical fusion device, should be centred on Nagoya University's Institute of Plasma Physics. But it has taken time for the fusion research community to reach a consensus on organization of the new institute and to persuade the

Plans for state slug salted away

Washington

CALIFORNIANS will have to do without a state mollusc. Governor George Deukmejian last week vetoed legislation that would have made the banana slug (*Ariolanax*) the official symbol of the order Mollusca in California.

In his veto message, Deukmejian explained that he did not believe the banana slug was indigenous to California, as it was first identified at the mouth of the Columbia River in Oregon. He also felt that the state mollusc — if there is to be one — should be more representative of the international reputation the state enjoys. Black abalone, pismo clam, common squid and red abalone were all mentioned in the veto message as examples of other molluscs that make a home in California. A move in the state Senate that would have substituted the red abalone for the banana slug failed.

Even without a state mollusc, Californians have a varied array of emblems from the animal and plant kingdoms. There is a state fish (California golden trout, *Salmo gairdneri*), bird (Valley Quail, *Lophortyx californicus*), insect (Dog-face butterfly, *Zerene eurydice*), tree (redwoods, both varieties, *Sequoia sempervirens* and *S. gigantea*), animal (grizzly bear, *Ursus horribilis californicus*), reptile (desert tortoise, *Gopherus agassizii*), marine mammal (grey whale, *Eschrichtius gibbosus*), flower (golden poppy, *Eschscholtzia californica*) and fossil (sabre-tooth cat, *Smilodon californicus*).

Joseph Palca

Ministry of Finance that a second major fusion machine is required.

The new institute will be at Toki in Gifu Prefecture. The ministry of education has spent about ¥2,500 million (\$20 million) over the past three years to buy and level the site — the land, although cheap by Japanese standards, was rather hilly.

And last week, the ministry, in budget requests for next fiscal year, applied for ¥572 million to begin construction of the institute and ¥2,480 million for research and development of the new helical machine. A decision on the budget requests will be made at the end of this year.

The new institute, which will open at a temporary site on the Nagoya University campus next April, will be for 'joint university use', like the Institute of Space and Astronautical Sciences in Tokyo and the High Energy Physics Laboratory in Tsukuba, and will draw about 150 researchers from Nagoya, Hiroshima and Kyoto universities. Kyoto University leads in the development of helical fusion devices, and Professor Atsuo Iiyoshi, head of the heliotron group at Kyoto, is chairman of the design group for the new machine.

Iiyoshi hopes that the institute will have a total staff, including technicians and administrators, of "300 or more" and at least ten posts for foreign researchers, when the buildings at Toki are completed in about three years.

The helical machine will have a major radius of 4–5 m and a magnetic energy of 2 GJ, by far the largest helical device in the world. The design group has opted for superconducting magnets which will require several years of research and development and the machine is not expected to begin operations before 1995.

Several hundred million dollars will be required to build the device, according to Iiyoshi, and industry is keen to win the contracts for the superconducting magnets and peripheral equipment, such as the supercomputer, he says.

So what will happen to the ministry's other fusion projects? The Institute of Laser Engineering at Osaka University is a world leader in laser fusion. The institute recently installed frequency converters in its huge Gekko XII laser to obtain 'blue' near-ultraviolet light which will allow experiments closer to break-even conditions. And Professor Sadao Nakai, director of the institute, hopes to upgrade the laser from 30 to 100 kJ to reach the break-even point. But with the ministry committed to the new helical device this project will probably have to wait.

David Swinbanks

NASA sets launch schedules for future missions

Washington

NASA, the US National Aeronautics and Space Administration's last week released the latest launch schedule for both the shuttle — now scheduled to resume operations either later this month or early next — and its fleet of expendable launch vehicles (ELVs).

Because of inflexible launch windows, the three forthcoming Solar System missions retain their prized manifest slots, but the Hubble Space Telescope's launch date has been pushed back by seven months. Priority in shuttle assignments went to missions with fixed launch windows and those which require the specific capabilities of the shuttle. Magellan, the

The delay will also mean that more telescope systems can be made ready before launch. Giacconi says hardware and software improvements have been proceeding smoothly. New solar cells provided by the European Space Agency will improve power levels by about 15 per cent, and switching from nickel-cadmium to nickel-hydrogen batteries will improve energy storage and cycling capabilities.

There has also been a compromise on the mission profile for the telescope. At first, when the launch was expected during a period of high solar activity, mission managers sought to place the telescope into the highest orbit possible so it could tolerate a more rapid decay. But NASA

NASA's launch schedule

Sep	88	ELV (Atlas)	NOAA-H
Sep	88	Shuttle (Discovery)	TDRS-C
Nov	88	Shuttle (Atlantis)	DoD
Feb	89	Shuttle (Discovery)	TDRS-D
Apr	89	Shuttle (Atlantis)	Magellan
May	89	ELV (Delta)	Cosmic Background Explorer
May	89	ELV (Atlas)	NOAA-D
Jul	89	Shuttle (Columbia)	DoD
Aug	89	Shuttle (Discovery)	DoD
Sep	89	ELV (Atlas Centaur)	FLTSATCOM-F8
Oct	89	Shuttle (Atlantis)	Galileo
Nov	89	Shuttle (Columbia)	LDEF-1R, SYCOM IV-5
Dec	89	Shuttle (Discovery)	DoD
Feb	90	Shuttle (Atlantis)	Hubble Space Telescope
Feb	90	ELV (Delta)	ROSAT
Feb	90	ELV (Scout)	Transit-27
Mar	90	Shuttle (Columbia)	Astro-1, BBXRT
Apr	90	Shuttle (Discovery)	Gamma Ray Observatory

Venus radar mapper, will be the first pure science mission to launch. The NOAA-H satellite — primarily an operational satellite for weather forecasting — has a solar back-scatter ultraviolet measuring instrument that can give vertical profiles of atmospheric ozone, but does not provide total atmospheric ozone measurements.

The Galileo mission to Jupiter and the Ulysses mission that swings out of the ecliptic at Jupiter to study the Sun's poles have essentially the same launch windows, but a decision was made last year to give Galileo the 1989 launch and Ulysses a 1990 launch.

It is the space telescope that has absorbed the biggest delay, but although the delay is a blow to morale, Riccardo Giacconi, director of the Space Telescope Science Institute, says he is more convinced than ever that NASA is committed to the mission, and will do everything possible to permit it to be launched under the current schedule. Giacconi says an example of the commitment is NASA's willingness to keep the telescope 'ready to launch' from August next year. This means that if any intervening payload is cancelled, the telescope can fly instead.

was unhappy with this, because if the expected decay did not occur, the shuttle would be at the edge of its performance capabilities when it attempts a planned repair mission. The repair mission could place more demands on the shuttle than deployment, as precise matching of orbits will require a large propellant margin. With deployment, slight deviations in projected orbits can easily be tolerated.

Under the compromise, the shuttle's main engine will boost the telescope into an elliptical orbit 310 by 330 nautical miles. The orbit can then be converted into a circular one at any altitude in that range depending on existing solar activity.

The delay did mean scheduling headaches for NASA managers who will have to transport the telescope from Sunnyvale, California, where it is stored in one of the world's largest clean rooms, to Cape Canaveral in Florida. The spacecraft was to be moved in November on a reinforced-hulled ship normally used for supplying arctic fleets, but that has been cancelled. Now NASA is looking into the possibility a flight aboard a C-5A jet transport, or possibly a later ocean ride.

Joseph Palca

Genentech patent

GENENTECH last week added yet another significant biotechnology patent to its portfolio. The California company announced that it had been granted a US patent covering the human recombinant version of the blood clot-dissolving drug tissue plasminogen activator (TPA) and the means for producing it.

The patent's 11 claims outline Genentech's ownership of the DNA sequences coding for human TPA, its exact amino acid sequence, and all recombinant vectors, microorganisms and cell lines which are used to produce it. TPA was first cloned and expressed in 1982 by Genentech employees together with Desire Collen from the University of Leuven in Belgium (see *Nature* 301, 214; 1983). A US patent covering purified TPA was licensed by Genentech in June (see *Nature* 333, 790; 1988).

The latest patent covers the same technology as Genentech's UK patent on TPA, which is currently the subject of a court battle with Wellcome plc (see *Nature* 328 189; 1987). Based on the new patent, Genentech has filed US suits against Wellcome plc's US subsidiary, Burroughs Wellcome, and its collaborator, Genetics Institute.

C.E.

New federal awards

THREE US federal agencies last week jointly announced new awards totalling \$9.5 million for plant science centres. The largest award — \$5.7 million over five years from the National Science Foundation (NSF) — goes to Cornell University for a centre for the experimental analysis and transfer of plant genes.

The three awards are part of a joint Plant Science Centers Program established by NSF, the Department of Agriculture (USDA) and the Department of Energy (DoE). USDA has awarded \$1.3 million to Michigan State University for a centre for genetic and biochemical alteration of plant lipids and starch, and DoE will give Arizona State University \$2.5 million for a centre for the study of early events of photosynthesis. NSF will add \$610,000 for equipment costs at the Arizona centre. J.P.

Collaboration begins

FOUR projects linking research institutes in East and West Germany were given the go-ahead on 22 August in East Berlin. The projects, which include nuclear reactor safety research and high energy physics, are among 27 joint projects agreed upon in principle by the two states during the visit of East German leader Erich Honecker to Bonn in September, 1987. Significantly, the new projects include researchers from West Berlin, whose participation had been the sticking point for many years in negotiations over scientific cooperation between West Germany and the Soviet bloc.

S.D.

Changing relationship seen in new corporate–university ties

Boston

Du Pont Electronics last week publicly unveiled a fledgling \$1.5 million university/industry programme with 12 institutions including the Massachusetts Institute of Technology (MIT), Syracuse University and the University of Arizona. A spokesman for the 'Research Alliance' programme proclaimed that Du Pont had "hired schools to work as an extension of the company's R&D staff", a description that displeases some academics.

Du Pont's programme involves approximately 15 individual projects ranging in scale from \$50,000 to \$150,000

Industry calls for review of teaching

London

IN the distribution of resources among UK universities, there should be less emphasis on the quality of research and more on the quality of teaching, according to a pressure group of university vice-chancellors and company chairmen, the Council for Industry and Higher Education. In a response last week to the exercise being carried out by the Universities Grants Committee (UGC) to evaluate research, the council says that industry values teaching above research, and it urges the UGC to carry out a review of teaching similar to that for research. In industry, there is "widespread and often passionate concern" about teaching, says the council, and a review "which seems to judge universities' resource-worthiness by their research prowess only, gives just the wrong emphasis".

But the UGC rejects the criticism outright. The review of research is only one of many assessments carried out by the UGC and teaching is evaluated, says a spokesman. "That criticism is idiotic."

The council also criticizes the research review itself. Although it accepts that the review is necessary, it stresses that it should not be based simply on the volume or value of a department's research contracts, as this would lead to a bias towards applied work which is easier to "sell", and companies do not support this shift.

It also warns that the review might produce a three-tier structure of universities which would be "unnecessarily rigid" and would damage morale and the quality of teaching. But the UGC says the criticisms are ill-conceived and invalid. Research is not evaluated solely on the merit of research contracts, and a three-tier structure may not emerge from the review process, says a spokesman.

Christine McGourty

and involving among other things optoelectronics, thermal conductivity and computer-aided design of electronic circuitry. The programme is an example of many similar new corporate/university arrangements which move away from open-ended 'no-strings-attached' research grants towards more directed research. Noting that US universities used to "remain aloof" from such directed research arrangements with corporations, Dr Rudy Carboni, director of Du Pont's Research Alliance, says the programme's existence is an indication of a "very, very big change in the climate at American universities".

With federal dollars accounting for less total university research spending, universities have turned increasingly to corporate support. At MIT, for example, Carol Van Akin, director of the Office of Sponsored Research, calls industrially funded research the school's "most significant growth area in funding". Industry funding at MIT has risen 20 per cent annually since 1976 and now totals about 15 per cent of all money spent on research at the school.

This has led to an increasingly entrepreneurial atmosphere in academic departments. Carboni and many others praise

the changes as a key strategy to bring products to market quickly in fast-paced, high-technology fields and keep the United States competitive with Japan and other countries. But others have reservations. Sheldon Krinsky, philosopher of science at Tufts University, believes that some of the new arrangements threaten to skew research priorities. Krinsky stresses the "different rules" that guide research in industry and academic life.

Du Pont's Research Alliance is just one of many such arrangements in the newly evolving relationship between industry and academic life in the United States. In late 1986, Washington University caused raised eyebrows when it became the first university to go into business, forming its own private marketing arm in a joint venture with a subsidiary of Monsanto. In the arrangement, the new corporate/university entity would help university researchers to commercialize and market biomedical and other high-technology products that came from their research.

Carboni says that Du Pont's programme is a more personalized effort than many previous arrangements in that the company tailors the projects directly to selected academic researchers. But Carboni acknowledges the highly directed nature of the research. "This research is not funded by an unrestricted grant", he notes. "Researchers have to stay within the agreed-upon objectives." **Seth Shulman**

Fault in Indian communications satellite not worth repairing

New Delhi

THE Indian Space Research Organization (ISRO) has decided to abandon its efforts to repair the fault in the INSAT-1C satellite which was launched on 21 July. The \$140-million satellite has been crippled by "a massive short circuit" in one of the two power buses. That means only half the planned power is available for operating the television, telecommunications and meteorological payloads.

For more than four weeks, scientists from ISRO and the American Ford Aero space Corporation (FAC), which built the spacecraft, have been carrying out simulations on the ground to find a fault clearance plan. It became evident that repairing INSAT-1C carried a risk of losing even the limited services the satellite can now provide. ISRO has thus found it "prudent" to keep the spacecraft in its present condition "rather than risk corrective measures that may or may not succeed".

According to space commission chairman Professor U. R. Rao, repairs will be attempted only if a crisis threatens the survival of the spacecraft. The satellite can at present handle 2,000 telephone calls (instead of 4,000) and one television pro-

gramme (instead of two). The power failure will not affect on the ability of the satellite to send pictures of cloud cover.

More serious, according to ISRO, is the problem of maintaining the thermal balance of the satellite consequent to the loss of one electrical bus—a problem that will be accentuated during the satellite eclipse seasons. The spacecraft has also lost the redundancies in vital subsystems such as telemetry, telecommand, and attitude and orbit control systems. Should any of these systems develop problems, the satellite will become useless.

According to ISRO, INSAT-1C has been insured for \$72.6 million at a premium of \$11.05 million. The insurance claims will be lodged after the stipulated period of 180 days.

It is not the first time ISRO has had problems with satellites manufactured by FAC. INSAT-1A failed within three months after launch in 1982, and its successor INSAT-1B had difficulties in unfurling the solar panels and the solar sail. These design defects were corrected in INSAT-1C but, according to Rao, no one expected it to develop a fault in its power supply. **K. S. Jayaraman**

Scandal over Soviet artificial blood research project

London

THE darker side of the Soviet Academy of Sciences was revealed last week with a claim by Dr Genrikh Ivanitskii that a major research project was closed down improperly on the initiative of the late Dr Yuri Ovcinnikov, for many years head of the Institute of Bioorganic Chemistry of the Soviet Academy of Sciences. The project concerned the search for a synthetic fluid that could temporarily serve as a replacement for blood.

According to Ivanitskii, who lost his job as director of the academy's Institute of Biophysics at Pushchino in the infighting, the research was stopped because a 'legend' had grown up in the academy that no serious work in medical biology could be carried out without the participation of Ovcinnikov and his institute.

According to Ivanitskii, the research produced a product, *perftoran* — a perfluorocarbon, which represented a major breakthrough in medical research and hence a threat to Ovcinnikov's monopoly. Not all Soviet biologists would agree with Ivanitskii on the merits of *perftoran*. What is perhaps more important, however, is his suggestion that research can be closed down without a proper peer

assessment by people who are neither competent to judge nor impartial.

Ivanitskii's account, in a letter to *Literaturnaya Gazeta*, stated that the research carried out under a directive of the praesidium of the Academy of Sciences, signed by Ovcinnikov himself, was entrusted to two academy institutes, the Institute of Biophysics and the Institute of Elemento-Organic Compounds. Three promising compounds were synthesized which, Ivanitskii said, gave the Soviet Union the lead over foreign teams working in this field. Yet from 1983 onwards, the academy's praesidium neglected to make the required annual assessment of interim results, in spite of frequent reminders from the researchers. In 1984, the research was "slowed down", and work on *perftoran* and other blood substitutes was omitted from the 1986-90 plan.

Furthermore, from 1982 onwards, the laboratory at the Institute of Biophysics where the *perftoran* research was going on became the subject of an investigation by the public prosecutor's office. The Serpukhov Town Party Bureau became highly critical of the work of the laboratory (which it was not competent to

judge). Several members of the team left the institute or transferred to other laboratories. Finally, in December 1985, the head of the team, Dr F. Beloyartzev, committed suicide, leaving a note that he could no longer bear to live "in conditions of calumny". A "crudely tendentious" posthumous attack on him appeared in *Sovetskaya Rossiya*, and the investigation into his suicide was carried out, Ivanitskii alleges, in a very one-sided manner.

When a commission was eventually convened by the academy to assess and wind up the research, it was headed by Academician Yakov Kolotyarkin, a physical chemist who had, he himself states, no knowledge either of medicine nor of the working of the Institute of Biophysics. This, according to the then president of the academy Dr Anatolii Aleksandrov, would keep things impartial. In fact, in spite of his lack of inside knowledge, Kolotyarkin soon realized that the commission was not impartial. The documents submitted to it contained "a mass of unanswered questions", and a suggestion from Academician Aleksandr Fokin, director of the Institute of Elemento-Organic Compounds, that the commission should include scientists who had worked on *perftoran* was vetoed by Ovcinnikov.

At the same time, the question of Ivanitskii's dismissal was raised, although, when Kolotyarkin challenged Ovcinnikov, the latter admitted that there had been no complaints about Ivanitskii's work. Nevertheless, while Kolotyarkin, the chairman of the commission, was ill in hospital, the other members closed down the research and dismissed Ivanitskii.

In addition to Ivanitskii's letter and Kolotyarkin's account of the working of the commission, *Literaturnaya Gazeta* also published testimonials to Beloyartzev, Ivanitskii and *perftoran* from Academician Fokin and from two doctors who had used *perftoran* in clinical trials. Their testimony adds up to a sharp attack on the late Dr Ovcinnikov — somewhat ironically in view of their rebukes to *Sovetskaya Rossiya* for attacking Beloyartzev when he was no longer alive to defend himself.

Literaturnaya Gazeta's main concern with what it calls the "banal and at the same time tragic history" of the *perftoran* affair is that, in spite of *perestroika* the danger of such incidents remains. Only recently, *Literaturnaya Gazeta's* commentator noted, *Izvestiya* reported a similar conflict between the Computer Centre at Pushchino and the Serpukhov Town Party Bureau (the same as had improperly criticized the Institute of Biophysics). The lesson of the *perftoran* affair, *Literaturnaya Gazeta* said, is that there should be some legal provisions to defend scientists and their work from arbitrary interference by administrators and bureaucrats — including academic ones.

Vera Rich

Industrialists in force on new funding body

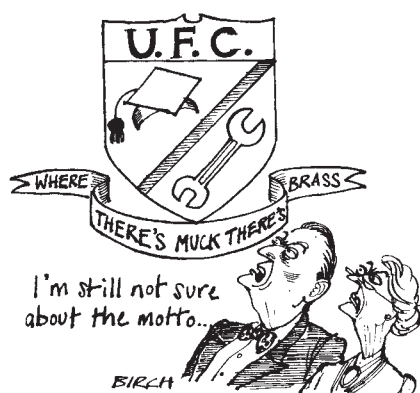
London

INDUSTRIALISTS are a significant force on the new Universities Funding Council (UFC) which next April will replace the Universities Grants Commission (UGC) as adviser to the UK government on grants to individual universities. The government announced the composition of the council last week. Four members are from industry, one is from commerce and two have interests both in industry and in the academic world, including the chairman of the council. Of the 14 members of the council, the number with business interests is seven, compared with only four of the UGC's 18 members. Sir Peter Swinnerton-Dyer, chairman of the UGC, becomes chief executive of the new body.

In some parts of the science community, there is concern that the strength of the industrial interests on the council will result in research being too closely directed towards the needs of industry at the expense of freedom to pursue curiosity-driven research. Dr John Mulvey of the University of Oxford, who is also secretary of Save British Science, says concern would be justified if criteria for supporting research were no longer primarily scientific. Otherwise, he says, the move is a sensible one which will improve industry's

understanding of universities.

Swinnerton-Dyer says concern over the criteria for funding research is "unreasonable" and that in the transition from the



UGC to the UFC, changes "are not going to be dramatic".

The UFC is perceived as having greater powers and more direct control over universities than its predecessor under the powers granted to it by the 1988 Education Reform Act (see *Nature* 334, 371; 1988) but the government denies this is so.

Christine McGourty

New set of eyes on the radio sky in the Southern Hemisphere

Narrabri, New South Wales

AFTER a decade of planning, design and construction, Australia has entered the big league in world astronomy with the commissioning last week of the Australia Telescope.

The telescope is a synthesis telescope, consisting of eight telescopes at three sites in New South Wales. Six of the antennas — comprising the Compact Array — are at the Paul Wild Observatory at Culgoora in north-west New South Wales. Each of the antennas is 22 m in diameter, and five are movable along a 3-km track. The sixth lies a further 3 km to the west. The two other antennas making up the Australia Telescope are a new 22-m diameter dish near Coonabarabran, and the 27-year-old, 64-m radio telescope at Parkes.

By itself, the Compact Array can map the same fine detail as a telescope 6 km in diameter. At the 3-cm wavelength it has a

resolution of 0.8 seconds of arc — equivalent to being able to read a telephone directory from 200 m away. When combined with the two other telescopes to make the Long Baseline Array, resolution increases 50-fold.

The Australia Telescope will be used with telescopes in Australia and in other countries including South Africa, Spain, the Soviet Union and the United States for very long baseline interferometry (VLBI). Australia will also participate in RadioAstron, a space VLBI project of the Soviet Space Agency scheduled for launch in 1992, and the US Quasat (quasar satellite) mission now scheduled for 1996.

The Australia Telescope will open new regions of the radio sky to astronomers. Current Northern Hemisphere telescopes are unable to look further than 37° S, providing only a limited view of the centre of our Galaxy.



Prime minister Hawke opens the Australia Telescope (AP).

US astronomers gain access to database

Paris

A NEW agreement between the French national information-technology and automation research institute (INRIA), in Nice, and two US government agencies, the National Science Foundation (NSF) and the National Aeronautics and Space Administration (NASA) will make it easier for US astronomers to gain access to a key astronomy database in Strasbourg. Until now, computer incompatibility and high costs have made this resource inaccessible to all but a few US astronomers.

The database, SIMBAD (set of identifications, measurements and bibliography for astronomers), enables astronomers to look up an astronomical object by its astronomical designation, gaining access to nearly all known information and a listing of papers about the object since 1950. But the US Department of Defense standard communications protocol, TCPIP, widely

used in American networks, is not compatible with the X.25 protocol used in Europe, making it difficult to exchange facilities across the Atlantic. The high costs of using the database, about \$100 per connection, was also beyond the reach of most astronomers.

Now, with funding from NSF and NASA, these obstacles have been overcome. NSF has paid for about 75 per cent of the costs of a permanent satellite network link between Princeton University and INRIA's Nice campus, with NASA financing the rest, as well as underwriting the charges for use of the database itself. Meanwhile, INRIA has developed a gateway (dedicated computer and software) to translate US and European protocols.

After some initial teething troubles when the link was demonstrated in early August, the system is now said by an INRIA spokesman to be fully functional. Peter Coles

The Commonwealth Scientific and Industrial Research Organization (CSIRO) Division of Radiophysics and private industry built the new telescope at a cost of A\$50 million. It will have an annual operating budget of A\$5 million. Dr Ronald Ekers, recently returned to Australia from the Very Large Array at Socorro, New Mexico, in the United States, will be the director.

"The Australia Telescope has a high dynamic range and a wide field of view", says Ekers, "as well as high angular resolution and sensitivity. Australia has now become one of the leaders in world astronomical research." **Tania Ewing**

West Germany drills deeper

Munich

WHY are West German researchers drilling a 14-km-deep hole in Bavaria? Are they seeking oil? Or gas? Not at all, explains Dieter Renz of the Research & Technology Ministry (BMFT). They simply want to see what there is to see. The BMFT-sponsored project is one of several 'super-deep drilling' projects around the world that will shed light on the structure of the Earth's crust.

Renz has just returned from an international conference at Jaroslavl in the Soviet Union where the Soviets revealed plans to drill 11 holes to depths ranging from 8,000 to 15,000 m. The occasion was only the second at which Westerners had been invited to visit a Soviet deep-drilling rig, this time at Kiluoi Rog near the Black Sea.

In the United States, a hole is being drilled near Cajon Pass in California near the San Andreas fault and has reached a depth of 5,000 m.

West German researchers successfully reached a depth of 2,970 m on 1 September with a pilot borehole at Windischeschenbach in Bavaria. They will drill down to 5,000 m using the same narrow bore, regularly extracting core samples, in preparation for drilling of the main hole, which will begin in late 1989.

Researchers hope the hole will tell them more about the mechanics of rock formation at depth. The pilot borehole has already revealed interesting results, according to Renz. The rock exposed so far is crystalline, composed primarily of gneiss with varying sillimanite, kyanite and muscovite content. The graphite content is surprisingly high, accounting for the high level of electrical conductivity observed from the surface.

The pilot hole is being drilled using a diamond-tipped bore six inches in diameter lubricated with a viscous silicate compound. Almost all the rock is recovered and analysed. **Steven Dickman**

Unreproducible results

SIR—Amid the palaver generated in both the scientific and the popular press by investigations into the unreproducible results reported by Weaver *et al.* (*Cell* 45, 247; 1986), and more recently by Davenas *et al.* (*Nature* 333, 816; 1988), nobody seems to have noticed how odd the tone of both investigations has been. The most logical explanation for the irreproducibility of both the Weaver *et al.* and Benveniste group results is simple human error: sloppy experimental procedure, or perhaps unforeseen artefacts that were not adequately controlled for, to be more charitable. Nobody who has done experimental science in a serious way would be surprised at that: it can be very difficult to design experiments so that all sources of systematic error are accounted for, and scientists, being human, make mistakes. Added to this fact is the common tendency of research workers to have a vested interest in a particular theory, which can cause them to overinterpret data or to be careless about considering alternative explanations, especially those that involve artefact. Yet the tone of the investigation into both papers has been that of a grand jury looking for evidence of a crime. The elaborate security precautions taken by the *Nature* team on the expedition to the Benveniste laboratory are inconsistent with a simple inquiry into human error. Rather, they are the precautions of people who expect to encounter deliberate falsification.

Scepticism is healthy in science, but cynicism can have unfortunate consequences. One of them is the extreme defensiveness that it creates in its targets. In his reply to the *Nature* investigation, Benveniste adopts a tone of the falsely accused that culminates in an anguished diatribe against "Salem witch hunts or McCarthy-like prosecution". This is not a reasonable response for a person who has only been shown to be in error (and that may be more revealing about its author than he realizes) but it is consistent with the tone of the times. And it seems to me that we are all served poorly by such an atmosphere. It polarizes discussions, creates an emotional rather than an intellectual climate for investigation, and leaves the public with a sense that science is a monolithic institution, hostile to new ideas and venomous in its defence of established ones.

What is the cause of this climate? I think it is the perception on the part of the researchers that to make an error is the worst thing that can happen to a scientist. Science is now a huge activity, with large numbers of laboratories competing intensely in almost every field. Frequently, the only way one's own discoveries can be distinguished from those of one's

competitors is by exaggerating the mistakes or wrong interpretations of the opponents.

When error can result in permanent harm to one's career, and when every mistake is seized upon as evidence of incompetence, a defensive attitude is inevitable. Excessive defensiveness leads to compounding of the error, for the researcher is then unwilling to believe evidence that a mistake was made. So pervasive has this process become that it is now considered harmful to one's reputation and career to publish what turns out to be a wrong interpretation of data even if the data are accurate and admit of more than one interpretation. The result, of course, is that people cling to incorrect interpretations longer than they should, which sets back the entire scientific process.

Fraud is a very serious matter, but I think it is more apt to occur in a climate where mistakes are treated too harshly. How many instances of deception actually begin as attempts to cover up honest mistakes I do not know, but surely when the consequences of making an honest mistake are nearly as severe as the consequence of fraud, and when the machinery used to investigate error comes to resemble that used to deal with suspected fraud, more people will consider the use of falsification to avoid the stigma of having

SIR—I submit that it would be easier to prove an incredible result like Benveniste's to scientists than it would be to disprove it to homoeopaths. Scientists accept the verdict of reproducibility. It is, of course, impossible to duplicate exactly any experiment. One must discount the effect of a myriad of minor variables. Homoeopaths, as evidenced by the comments of David Reilly (*New Scientist* 4 August 1988, p.30) are not willing to do this. If Benveniste's experiment turns out to be irreproducible, he says, diurnal rhythms or some electrical pollution in the laboratory may be the cause. Within the context of homoeopathy, this might make sense. Who is to say that leaving a television on next to the laboratory is unimportant, or that the experiment must be done at night, or not within 100 km of a nuclear power plant and so on? PETER J. LIPOWICZ
716 Turner NE,
Albuquerque, New Mexico 87123 USA

SIR—In my teaching, I encourage students to review freshman chemistry. Many resist and some insist that as biologists they do not need to know chemistry. If they intend to pursue imitative research, such indifference may be defensible. On the other hand, original research requires a general science background suf-

ferred. Thus, the present climate may actually increase the likelihood of the offence it is trying to stamp out.

What is needed is a decriminalization of error. Science often advances on the strength of theories that turn out to be incorrect, for a wrong hypothesis can produce many excellent experiments. (Columbus, after all, acting on a false hypothesis, bumped into a new world.) Appointments committees should look at the total record of a scientist, and not overemphasize a single mistake, especially if the mistake was corrected by the person who made it. University professors who train graduate students and postdocs must adopt this attitude and teach it to their students, so that the next generation of scientists will start from a different ethos.

I do not believe the decriminalization of error will lead to more error. Mistakes happen because careful research is hard and fallible human beings try to do it. But a climate in which self-correction is encouraged makes detection and admission of error easier. And it is the detection of error, and its correction, that is the real problem. All that is needed is the recognition, commonplace in other forms of human endeavour, that mistakes are the path to growth, especially when one discovers one's own, and learns from them.

GREGORY A. PETSKO

Department of Chemistry,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

ficient to recognize nonsense. The 'controversial' basophil neutralization paper nicely illustrates the distress that can ensue when a biologist knows less of atomic theory and water structure than a competent freshman chemist. Whatever embarrassment accrues to the authors of this paper should serve as a warning to those who promote lofty principles from narrow understanding. Perhaps narrow technical interests should be rewarded with special degrees such as Master of Mediocrity or Doctor of Imitation (D.Im.).

LESLIE C. LANE

Department of Plant Pathology,
University of Nebraska,
Lincoln, Nebraska 68583-0722, USA

SIR—As Randi so rightly says, reports of unicorns need to be checked with particular care. Readers of *Nature*, and past and prospective authors, deserve an explanation of how it was that *Nature* did not make the elementary check to see that Benveniste's degranulation counts were consistent with the expected poissonian distribution, either when he first submitted the paper, or before the extended paper was accepted for publication.

B.H. BLAND

Building Research Station,
Garston, Watford WD2 7JR, UK

Pugwash changes guard and style

The 38th annual meeting of Pugwash last week was unusually decorous but may have been nonetheless valuable on that account, if only because the organization is seriously pondering its future.

Dagomys (Caucasus)

THE Pugwash organization, unaccustomed to high living, last week set the world to rights for another year in the lap of southern Soviet seaside luxury in a 24-storey hotel of perplexing design. It also laid the foundations for a succession of the generations in its own management, listened with great attention to Andrei Sakharov, brooded more explicitly than usual about its future role and narrowly evaded a hijack by environmentalists.

The 38th annual meeting, larger than usual with more than 200 participants, was the first to be held in the Soviet Union since the election of Mr Mikhail Gorbachev as general secretary of the Communist Party. Manifestations of *glasnost* abounded (although Western newspapers are still mostly unavailable): people were repeatedly diverted from their special interests in arms control by the willingness of the sprinkling of their Soviet hosts to gossip about the changing social and political scene: Pugwash participants, at least, are learning that criticism of particular government decisions does not imply disloyalty.

Few Soviets

Soviet participation in the group has been changed. Academician V. Goldanskii is the new chairman of the Soviet committee; his fellow Soviet members on the international Pugwash council are the liberals Sergei Kapitzin and Anatoly Gromyko (whose Institute of African Studies provided the foreign-speaking guides and stewards for last week's events). The new Soviet committee's success in raising 850,000 roubles (nominally \$1.35 million) to support this year's meeting has dismayed its US and British counterparts, responsible for the next two meetings, who suspect they will not do as well. Soviet participation this year, while distinguished, was sparse both because of the novelty of the Soviet organization and, some guess, because the Soviet Union, with its many commitments in the field, is running short of knowledgeable people.

One theme informing discussion of Pugwash's traditional concern with arms control is the widespread opinion that the recent relaxation of tension between East and West is both an opportunity (for arms control agreements going beyond last year's treaty to ban missiles of intermediate range from Europe) and a cause for

urgency (because the relaxation may not last). Another is that the concept of national security is more than a military matter (but nobody actually said that security is too important to be delegated to generals). What follows is a gleaning from a week-long torrent of proposals.

Pugwash promises to become an even more vigorous advocate of 'non-provocative defence' in Europe, partly to sidestep the difficulties of enumerating and verifying conventional weapons systems of different kinds otherwise certain to arise in the negotiations on conventional European arms control in November, which will succeed the unproductive decade-long MBFR negotiations. (Battle tanks are bad, land-mines good, is an over-simple mnemonic.) But only little thought seems so far to have been given to the objections the military are likely to raise, although obdurate French chauvinism is widely condemned. The argument that tactical nuclear weapons in Europe should be abolished did not sweep everyone before it, but Sakharov advocated a 300-km demilitarized zone along the European frontier.

Suggestions for surmounting outstanding obstacles at the bilateral Geneva negotiations on strategic weapons, rehearsed in detail last week, include a ban on sea-launched cruise missiles (for fear that the verification of a ban on nuclear-armed cruise missiles will be impractical) and several ingenious schemes for counting the military potential of aircraft carrying cruise missiles in the balance sheets of warheads and delivery systems provisionally agreed at Geneva. Warhead reductions going beyond the 50 per cent cut now planned would require both an agreement not to deploy anti-ballistic missile systems and a comprehensive test-ban.

Optimism on the latter score, sustained by circumstances such as the fact that 40 US technicians are now drilling a vertical shaft near the Soviet testing-site at Semipalatinsk in preparation for a planned Soviet test a few weeks from now, is tinged with tactical disagreement. Some would prefer an interim agreement to reduce the threshold of the present treaty limiting the power of underground nuclear explosions from 150,000 tonnes of TNT equivalent to 10,000 or less, on the grounds that while the seismic detection of much smaller explosions is feasible, the discrimination from small or distant earthquakes, in the absence of more data on the seismic properties of Soviet surface rocks, would yield

too great an annual harvest of disputes. Others would go the whole hog now, relying on the reluctance of governments to risk being found in violation. Still others would prefer to set their sights higher, on a treaty to eliminate nuclear weapons by 2010 (advocated by Gorbachev but also the government of India), partly from anxiety that the Nuclear Non-Proliferation Treaty will unravel.

Enthusiasm

Enthusiasm for an international organization for military observation by Earth satellite mounts from year to year.

On chemical weapons, Pugwash seems united in condemnation of Iraq for its use of chemical weapons in the Gulf War and in support for the new convention being negotiated at Geneva to ban the production of military gases. But there are two opinions on verification: one is that verification is feasible now, and should not further delay signature, another is that there needs to be a research programme aimed at the development of techniques and instruments therefor. (Finland won praise for its development programme, which has sensors mounted in commercial aircraft to detect trace gases from chemical plants.) Last week saw the beginning of drum-beating in the cause of the unilateral renunciation of chemical weapons.

This year's meeting also broke new ground by the attention paid to military research and development "in driving the arms race", in part by creating a "demand for new weapons systems". Pugwash acknowledges that its influence is best exercised on the political decision-making from which research springs, but would have major military powers close to at least one military laboratory to reduce the feedback of the "self-serving" influence of military research and development on political decisions. Universities should be demilitarized, more attention should be paid to emerging fields (such as anti-matter, biotechnology, isotope separation by laser and even neural networks) with commercial and military applications.

The recognition that research controls are most effective when applied to the testing of weapons (for example nuclear bombs and anti-satellite weapons) did not prevent some from advocating agreements not to conduct academic research with military implications, general pessimism about the openness of all research notwithstanding. These issues are most

pointed in relation to Pugwash's advocacy of further ('deep') cuts of strategic arsenals, for which purpose restraints on space-based anti-missile systems are reckoned to be essential.

The importance separately attached to environment and development (of developing countries) this year was justified by the doctrine that security is not simply a military matter, by interest in and various degrees of anxiety about the greenhouse effect and, no doubt, by a wish to insure against the possibility that Pugwash has worked itself out of a job with the relaxation of East-West tension.

The upshot, a document called a manifesto and said by its authors to have been "crafted" rather than drafted, came near to causing a constitutional crisis. Headed "Insuring (*sic*) the survival of civilization", the document says that environmental degradation, "already a fact . . . can quickly lead to the destruction of civilization", and that "the present inequitable world economic order" restricts developing countries to environmentally destructive practices which, coupled with "world-wide population growth, excessive production, and profligate consumerism, will push the planet towards a global collapse".

Most of those speaking on the text last week were eloquently in favour, with the consequence that Academician Goldanskii embarrassed the new management of which he is a part by being the one to urge that the assembly should, after a vote, adopt the manifesto as a declaration of its opinion and to rank with the Russell-Einstein manifesto from which Pugwash sprang 33 years ago. But, sadly, Pugwash has learned the hard way that resolutions adopted in haste are usually regretted afterwards, and does not allow them. As one mildly dissenting participant pointed out, there is no assurance, as asserted, that "nuclear war is increasingly unlikely".

Sakharov

Sakharov's contribution to this discussion was tangential — put nuclear power stations underground, provide developing countries in which tropical forests are sited with money incentives not to cut them down (an idea attributed to his wife, Yelena Bonner) and take steps to "safeguard the genetic fund of man", now threatened with "violation by anthropogenism [the interpreter's word]" and by the stockpiling of "negative mutations".

How far, and how, Pugwash moves down this road remains to be seen. The case for embracing environment and development follows from the logic of its case that security is not merely military, but some old hands are fearful that the environmentalists will make uncongenial bedfellows. A flat declaration by an official of the World Wildlife Fund, after an account by Academician Roald Sagdeev

of cooperative space research, that it is "unacceptable" that there should be "all these projects, involving so much money and brains" while there are "so many people in misery and . . . our biosphere is at such risk" may have been quickly rejected by Victor Weisskopf (with a version of the question "What is survival for?"), but the bad manners had many participants sucking their teeth.

A new role for Pugwash is not, in any case, everybody's goal, although it is generally agreed that times have changed. At the outset, Pugwash was a unique means of communication between East and West, but now there are several. The organization's concentration on the dangers of nuclear weapons and the need for arms control may similarly have been distinctive in the first decade or so, but now there are many other organizations working in the field.

Officials, both past and future, also point to the way in which Pugwash has influenced events and even governments. Dr Martin Kaplan, secretary-general until the end of 1988, claimed last week a direct influence on the negotiation of the atmospheric Test-Ban Treaty (1983), the Non-Proliferation Treaty (1968) and the Anti-Ballistic Missile Treaty (1972) as well as other important and more recent developments. The claim is supported by Professor Francesco Calogero, the Italian theoretical physicist who takes over from Kaplan on 1 January next.

Perhaps sadly, the claim rests primarily on anecdotal and necessarily partisan evidence (which does not imply that it is untrue). Another part of the difficulty is that, in 38 years, Pugwash has made so many suggestions that one of them will be found to coincide with almost any beneficial development there may now be. Unfortunately, the time (in the 1970s) seems to have gone when Western governments encouraged knowledgeable officials to participate in annual meetings, if only to stiffen Pugwash participants' knowledge of their opposition. Yet it remains the case, by the testimony of several last week, that both the Soviet and West German governments now have a vivid interest in Pugwash proceedings, perhaps to ensure that interesting ideas do not go begging. (In an uncompleted discussion on whether Pugwash should aim at idealism or realism, one participant complained that Mr Gorbachev has stolen Pugwash's reputation for idealism.)

At least one delegation (the Polish) is mostly sceptical. One member said he had no direct evidence that Pugwash proceedings had changed the way in which governments had tackled problems in international relations or arms control. He was also one of those complaining that last week's meeting was too big to provide pointed discussion comparable in quality with those of the specialized workshops on

particular topics (now being held roughly once a month) and that too many of the participants have come newly (and naively) to the topics they discuss.

The contrary view, not necessarily inconsistent, is that the annual meetings are not merely working conferences but ways of stimulating public interest and of inducting newcomers into what is necessarily a kind of club. Interestingly, one newcomer was complaining that, these days, the proceedings of negotiations such as those of the strategic arms treaty at Geneva have become so complicated that only young people have the energy to follow them in detail.

Problems

The constitutional problem may eventually prove more taxing. International Pugwash has no members, but consists merely of its committees (whose members are effectively nominated by national committees) and two offices (in London and Geneva). Nominations appear to be by a kind of osmosis, the rules for attendance at annual meetings mystify many of those chosen and, while there are said to be central financial accounts, they are not published. The demand for more "democracy", raised publicly by one younger participant last week, does not carry the full force of logic while clubs are not forbidden by the law to exist, but more explicit rules of management would probably help Pugwash towards its goal of raising \$5 million for the foundation set up in Switzerland with the backing of Kenneth Kaunda, Robert McNamara and Helmut Schmidt.

Newcomers are also often mystified by the proceedings of the annual Pugwash meetings, with their interdiction on voting and the convention that even the reports over which working-groups struggle for hours do not themselves form part of the council's final declaration, which is a bowdlerized version of them. There are several explanations. In the old days, particularly, some delegations felt it necessary to return with particular forms of words, but the council also assumes to itself responsibility for continuity (from one year to the next) and for making sure that what it says publicly will command respect and, also, attention. Old hands, familiar with the process, seem content, but others chafe at the idea that their ringing calls for action or for changed attitudes (as in "we must learn to think in a new way" from one of this year's documents) do not hit the headlines.

For the rest, last week was the last public appearance as an official of Dr Dorothy Hodgkin as president. The plan is that she will be succeeded in that role next year by Dr Joseph Rotblat, for all practical purposes "Mr Pugwash", whose claim on his club's affections has not been soured in 38 years.

John Maddox

Bacterial genetics

A unicorn in the garden

Franklin W. Stahl

IN 1943, Salvador Luria and Max Delbrück demonstrated (*Genetics* 28, 491–511) that a strongly selective environment alters the genetic composition of a bacterial population by allowing the growth only of adapted mutants pre-existing in the population. Their experiment gave no support to the view that cells can mutate adaptively in response to a selective agent. Luria's and Delbrück's demonstration that bacterial evolution proceeds by natural selection of pre-existing mutants legitimized the genetics of these microorganisms. On page 142 of this issue, John Cairns, Julie Overbaugh and Stephan Miller challenge these credentials. They offer evidence that individual bacteria can, indeed, mutate adaptively in response to a selective agent.

Cairns *et al.* point out that Luria and Delbrück were unfair: they did not let *Escherichia coli* show us what it could do if given a decent chance. As a selective agent, Luria and Delbrück used bacteriophage T1, which instantly kills sensitive cells. By contrast, Cairns *et al.* challenge *E. coli* with an environment that simply fails to support the proliferation of the standard type in the culture, but does support growth of the mutant type. In the first of several examples, the standard type is a strain unable to use lactose as an energy source because of a chain-terminating codon in the structural gene for β -galactosidase.

Following the protocol of Luria and Delbrück, Cairns *et al.* grew parallel cultures of standard-type cells in the absence of the selective agent, lactose. They then plated the entire contents of each culture tube onto petri dishes in which the agar contains lactose as sole energy source. Mutation from Lac^- to Lac^+ was required for any given cell to give rise to a colony. If the Lac^+ mutants arose solely during growth in the culture tubes, lacking lactose, the plate-to-plate distribution of colonies would be clonal — mutations occurring in various generations in the different culture tubes would give rise to mutant clones containing widely divergent numbers of Lac^+ cells. On the other hand, colonies arising as a result of mutations to Lac^+ occurring after the cells were plated would be randomly distributed among the petri plates.

In the event, Cairns *et al.* report composite distributions, part clonal part random, indicating the occurrence of both classes of mutation to Lac^+ . By various strategies, they then direct attention to the mutations that occur on the plates. They show that these mutations to Lac^+ happen

only if lactose is present on the plates. Mutations to other phenotypes, which confer no selective advantage, do not occur. Additional experiments further the iconoclastic view that bacteria have mechanisms for making just those mutations that adapt the cell to an available energy source.

What's up? Can bacteria really direct their mutational processes? Did bacteria discover 'directed mutagenesis' before the genetic engineers did? Cairns *et al.* appear to eschew such an idea. (However, when I read "... bacteria are in the presence of lactose and under strong selection pressure to become Lac^+ " (my italics), I wonder whether they may feel a bit drawn to the notion.)

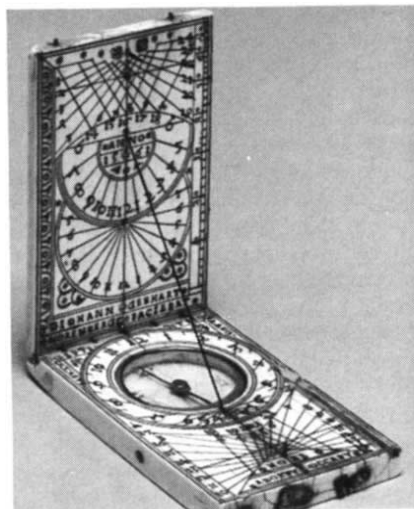
More attractive, the authors suggest, is some sort of trial-and-error mechanism that allows the selection of serendipitous molecular misconceptions that have an adaptive phenotype, with the selection operating at the molecular level rather

than at the level of the cell. How might we seek the mutable intracellular entities upon which selection can act to adapt an individual cell? Cairns *et al.* clearly indicate the sorts of models that might be considered. Readers will find it good exercise to try to build more highly specified schemes. Top marks will go to those who can build their model solely from familiar elements.

Here's one: my proposal is based on the post-Luria-Delbrück demonstration that mutation proceeds through a reversible intermediate. Polymerase-catalysed DNA synthesis has a mistake rate that is thousands of times higher than the observed mutation rate of freely growing cells. The fidelity normally observed is achieved by the action of post-replicative mismatch-correction enzymes. These enzymes remove stretches of newly synthesized chains whose sequences fail to be fully complementary to the templates on which they were made. To a good approximation, we can say that mutations result only when these correction systems fail.

I imagine that the Lac^- cells on lactose medium, starved by their inability to use the lactose, maintain some metabolism by cannibalizing macromolecules for energy

Exhibition of ivory-pocket sundials



IN THE sixteenth century, Nuremberg in south Germany became one of the most important centres in Europe for manufacturing ornamental goods and scientific instruments. This conjunction is strikingly illustrated by the numerous ivory pocket sundials produced by specialist craftsmen. Sixty of these beautiful objects can be seen in an exhibition at the Whipple Museum of the History of Science in Cambridge, UK, until 9 December 1988. Some of the sundials are small and merely functional but many are decorative objects that were proudly displayed by their owners. They tell the time by several systems, most of which were already obsolete when the dials were made, and provide astronomical and calendrical data of little practical use. Some carry extravagant tables of latitudes, used to adjust the dial for the various places the owner might imagine finding himself. A book and catalogue by Penelope Gouk explain in detail how the dials work and reconstruct their origins in the Nuremberg community of astronomers and craftsmen. The exhibition is open on Mondays to Fridays, and on Sundays during October, from 2–4 p.m. Left, a sundial by Johann Gebhart, made in 1561; right, detail from a sundial by Paul Reinmann, dated 1599, showing Judith beheading the Assyrian commander Holofernes. □

and 'building blocks'. In addition, they adventitiously replicate DNA segments, provoked (and primed) by the nicks and gaps of their outrageous fortune. In the lucky event that a synthesized DNA segment is adaptively mutant (on the transcribed chain), the resulting ability to utilize the lactose stabilizes the mutation by allowing the cell to replicate its chromosome. Now, let the starving *E. coli* adjust the relative rates of the several steps of mutation, replication and mismatch correction, and away she goes.

If you don't like my explanation, make one of your own after you read the paper of Cairns *et al.* You should be warned, however, that more difficult challenges are just over the horizon. A paper entitled "Adaptive evolution that requires multiple spontaneous mutations. 1. Mutations that involve an insertion sequence", by

Barry G. Hall, will soon appear in the journal *Genetics*. The phenomena reported there may require that bacterial cells can adapt by a sequence of two mutations, the first of which occurs at exalted frequency in the presence of the selective environment even though it offers no growth advantage by itself.

Widespread recognition of the claims by Cairns *et al.* may have to await experimental support for some explanation that makes use of familiar processes or establishes the biochemical basis of some unfamiliar ones. In the pursuit of same, as Hall points out, we are certain to uncover a wonderland of mechanisms by which bacteria cope with the harsh realities of the real world. □

Franklin W. Stahl is in the Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA.

Oxide superconductors

Magnetic theories in a spin

N. d'Ambrumenil

ELECTRON-SPIN excitations are thought by many to hold the key to superconductivity in the high-transition-temperature copper oxides. Yet the family of superconductors $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$, which shares many features with the cuprate materials¹, is reported by Uemura *et al.* on page 151 of this issue² to show no sign of static magnetic order (spin order) at temperatures down to 10 K. If there are no magnetic moments at all then this might imply that different mechanisms drive superconductivity in the two families of compounds.

$\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ is similar structurally to the high-temperature cuprate superconductors. The bismuth ions sit at the centres of O^{2-} octahedra rather like the copper ions in $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$. But in the bismuth compound the octahedra form a fully three-dimensional perovskite structure in contrast to the layered structures of the cuprates.

Strong evidence against a magnetic origin for superconductivity in the bismuth system is provided by recent neutron-scattering data analysed by Hinks *et al.*³ These show that the superconducting/normal phase boundary coincides with a structural phase transition. $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ moves from an almost cubic, superconducting phase for $x > 0.25$ to a tetragonal one, which is not a superconductor, for $x < 0.25$. The maximum value for T_c , the onset temperature for superconductivity, occurs for compositions close to the structural phase transition. As pointed out by Hinks *et al.*³, this is similar to the more conventional 'Chevrel-phase' superconductors, in which superconductivity is generally attributed to electron-phonon interactions. (Chevrel phases are molybdenum

sulphides based on the formula unit $\text{M}^{2+}\text{Mo}_6\text{S}_8$, with M any of Pb, Eu, Sn, Yb and Ba. Chevrel phases also have a structural phase transition and T_c is a maximum close to this transition.)

The structural phase transition in $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ is related to a change of the formal charge on the bismuth ions. In the tetragonal (non-superconducting) phase, bismuth exists in two different formal charge states: Bi^{3+} and Bi^{5+} . The existence of the two formal charge states implies an important role for charge transfer excitations associated with the disproportionation $\text{Bi}^{3+}\text{Bi}^{5+} \rightarrow 2\text{Bi}^{4+}$ as well as large distortions of the oxygen octahedra surrounding the bismuth ions. These features are the mainstay for theories of superconductivity based on the bipolaron and excitonic mechanisms. In the bipolaron model, two holes (electron vacancies) bind to a large lattice distortion to form a mobile quasiparticle⁴, whereas in excitonic mechanisms, holes are bound through electron-hole excitations. Both these mechanisms have been proposed for theories of the high-temperature cuprate superconductors, and in a recent News and Views article⁵, Rice tentatively identified $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ as the long-sought bipolaron superconductor. At this stage, the obvious hypothesis is that the superconductivity in $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ is related to the structural phase transition and (probably) the charge-transfer excitations.

The standard pictures of superconductivity in $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ and the other cuprates involve the magnetic moments on the copper ions playing an important role. LaCuO_4 , the parent or undoped compound, is a 'Mott' insulator in which

the Cu^{2+} ions carry a magnetic moment. These order antiferromagnetically at low temperatures. On doping with barium, holes are introduced into the Cu-O planes. The holes are thought to occupy orbitals predominantly on the oxygen sites^{5,6}. They also introduce defects into the antiferromagnetic order on the copper ions, eventually pushing the system into a spin-glass-like phase^{7,8}, but a phase in which the moments on the copper ions are still well defined. On further doping, the system becomes a superconductor.

Some theories of the superconducting phase treat the copper spins classically. These assume that the defects in the magnetic order induced by doping 'link' together so that the holes are no longer associated with any one defect and propagate freely (A. D. Jackson *et al.*, preprint). In this picture, one would expect a static or slowly varying magnetic moment on the copper sites. But most theories claim that the magnetic order vanishes completely in the superconducting phase, preferring instead to condense into a ground state which is a quantum spin liquid, sometimes termed the resonant-valence-bond (RVB) phase. In the spin-liquid phase there would be no static magnetic moment on the copper sites⁹. In all these theories pairing of charge carriers in the superconducting phase is closely associated with the spin of the copper ions.

So what are the magnetic properties of $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$? As already mentioned, in the non-superconducting phase $x < 0.25$, the bismuth ions exist in the two distinct formal charge states Bi^{3+} and Bi^{5+} . In neither of these states do the bismuth ions carry a magnetic moment as they are spin-zero configurations. This ensures the absence of any static magnetic order in the non-superconducting phase of the kind found in La_2CuO_4 . But the formal charge order could be thought to be analogous to the classical spin order in the cuprates. On doping (increasing x) holes are introduced in the Bi-O network and, if they occupy predominantly oxygen orbitals, they suppress the formal charge order as they do the spin order in the lanthanum copper oxides. The analogy could presumably be pursued in the superconducting phase, taking the charge carriers to be the oxygen holes, and might be the basis for a common theory of the two families of superconductors.

In the superconducting cubic phase bismuth exists in the formal charge state Bi^{4+} . In other words, all the bismuth sites are equivalent. As an isolated ion, Bi^{4+} has an $\text{spin-}\frac{1}{2}$ ground-state configuration and so technically could carry a magnetic moment in the solid although, as Rice pointed out⁴, a coherent superposition of the charge states Bi^{3+} and Bi^{5+} would also make the bismuth sites all equivalent. Magnetic susceptibility measurements

reported by Cava *et al.*¹ show $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ to be paramagnetic (ion sites having zero average spin) in the normal phase, the magnitude of susceptibility being comparable to that found in the cuprate superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. If a magnetic mechanism is to work for the yttrium compound it might still do so in the bismuth compound, although the absence of static magnetic moments reported by Uemura *et al.*² would appear to rule out theories which treat the moments classically, if not the novel spin-liquid phase³. It is also thought less likely, however, that the spin-liquid phase is the ground state in

a fully three-dimensional system than in the (effectively) low-dimensional cuprates with their Cu-O planes. □

1. Cava, R.J. *et al.* *Nature* **332**, 814–816 (1988).
2. Uemura, Y.J. *et al.* *Nature* **335**, 151–152 (1988).
3. Hinks, D.G. *et al.* *Nature* **333**, 836–838 (1988).
4. Rice, T.M. *Nature* **332**, 780 (1988).
5. Emery, V.J. *Phys. Rev. Lett.* **58**, 2794–2797 (1987).
6. Takahashi, T. *et al.* *Nature* **334**, 691–692 (1988).
7. Kastner, M.A. *Proc. Int. Cent. Theor. Phys. Towards the Theoretical Understanding of High T_c Superconductors*, 26–29 July 1988, Trieste.
8. Sinha, S.K. *Mater. Res. Soc. Bull.* 24–27 (24 June 1988).
9. Wheatley, J.M., Hsu, T.C. & Anderson, P.W. *Nature* **333**, 121 (1988).

N. d'Ambrumenil is in the Department of Physics, University of Warwick, Coventry CV4 7AL, UK.

Protein modification

Sticky fingers and CAAX boxes

Tony Magee and Michael Hanley

THE covalent modification of proteins is now recognized as the final common currency of signal transmission, protein trafficking to intracellular sites and metabolic regulation. There is, therefore, much interest in the widespread occurrence of post-translational addition of lipid-derived hydrophobic moieties to proteins. One class of hydrophobic modification is the attachment of proteins to glycosyl phosphatidyl-inositol (GPI), which acts as an anchor or 'greasy foot' for cell-surface attachment¹. Another class, the direct attachment to proteins of long-chain fatty acids, or 'sticky fingers', is thought to promote association of proteins with endomembranes. Two types of such modification (acylation) are known to occur in eukaryotic cells: amide linkage of exclusively myristic acid (C 14), and (thio)-ester linkage of (mainly) palmitic acid (C 16). A survey of acylated proteins (see table) emphasizes the diversity of substrates for these types of modification. The sequence requirements for myristoylation and palmitoylation are now being elucidated, potentially providing a predictive basis for the identification of new acylated proteins, as well as clues as to how these events control protein localization and function.

Myristoylation first generated interest when it was shown that the oncoprotein tyrosine kinase, pp60^{src}, requires myristic acid modification for its proper membrane localization and for its transforming activity². The sequences specifying myristoylation have been defined as residing in the seven amino acids at the amino termini of target proteins^{3,4}, from which a candidate consensus sequence can be proposed (see table). Glycine is invariably the donor of the α -amino group to the amide linkage with myristate, and other residues, particularly in positions 2 and 5 relative to the glycine at position 1, either enhance or reduce the efficiency

of myristoyltransferase action. Paradoxically, this modification does not always lead to stable membrane association, as many myristoylated proteins are found in the cytosol. Thus, myristic acid may, in some instances, cause a transient association with membranes as has been suggested for the '80K' protein kinase C substrate of macrophages⁵. The reversible association of a modified protein with membranes could provide a mechanism for some forms of translocation that are common in different kinds of cell activation.

Little is known about the sequence requirements for palmitoylation of certain types of substrate proteins, such as transmembrane receptors (which almost invariably occurs on a cysteine residue near the transmembrane/cytoplasmic interface, with basic residues nearby). But palmitoylation, together with a series of other post-translational modifications, has been extensively examined in another family of oncoproteins, the GTP-

binding *ras* and related proteins. The best-studied examples in the *ras* family have a conserved sequence CAAX (where C is Cys; A, aliphatic; and X, any amino acid) at the carboxy terminus which is required for acylation, for binding to the inner surface of the plasma membrane and for full biological activity, including transformation⁶. The cysteine residue has been thought to be the site of palmitic-acid attachment and is also required for an essential earlier processing event which in yeast is under the control of the *dpr1/* *RAM* gene^{6,7}. This processing event seems to involve removal of the last three amino acids, thus placing the cysteine at the carboxy terminus, where it can be acylated.

Now, yet another modification of these proteins has been identified: carboxymethylation, probably of the α -carboxyl of the carboxy-terminal cysteine⁸. Hence, this short region of the *ras* protein family is extraordinarily crowded with post-translational modifications. Unlike myristoylation, which is thought to be an essentially permanent modification occurring co-translationally, palmitoylation may be reversible. The palmitic acid on p21^{N-ras}, for example, turns over much more rapidly than does the protein itself⁹. Removal of palmitic acid may thus afford a means of releasing proteins in response to a stimulus, leading to modulation of their biological activities. This circumstance provides a direct intracellular parallel to the potential physiological removal of the extracellular GPI anchor, which could be cleaved to release surface proteins¹. Both myristic and palmitic acids may eventually be shown to confer a time-dependent, and potentially regulated, form of membrane attachment, albeit by different mechanisms.

A search of protein sequences reveals other interesting candidates for carboxy-terminal palmitoylation which contain the CAAX box⁸. These include the

Survey of acylated proteins

	Q N Consensus M G A T/S X S X signal* D D P F Y
Myristoylated proteins	
Oncoproteins	pp60 ^{src} , p56 ^{lck} , p120 ^{abl}
Signal-transducing proteins	PKA C-subunit, calcineurin B
G proteins	α_1 , α_0
Viral structural proteins	<i>gag</i> (mammalian retroviruses), VP4 (picornaviruses), p28 ^{orf} (HIV)
Palmitoylated proteins	
Viral fusion proteins	influenza HA, VSV G
Cell-surface receptors	transferrin-R, IGF-1-R
Oncoproteins	<i>ras</i> protein family
GTP-binding proteins	YPT1
Mating factors	yeast α -factor
Cytoskeletal proteins	ankyrin

* The single-letter amino-acid code is used. X, any amino acid. Residues above the line are tolerated, residues below are inhibitory. PKA, cAMP dependent protein kinase; VP4, viral polypeptide 4; VSV, vesicular stomatitis virus; IGF, insulin-like growth factor; YPT, a *ras*-related protein named after the *ypt* mutation in yeast.

α -subunits and one γ -subunit of several G proteins (α_o , α_i , α_s , γ) and the nuclear lamins. Reversible acylation of lamins might control their binding to the nuclear membrane during mitosis. It is interesting that lamin B, which remains tightly associated with the nuclear membrane throughout the cell cycle, appears to be modified with a different hydrophobic moiety, possibly an isoprenoid chain derived from mevalonic acid¹⁰.

The best-described examples of palmitoylated proteins are associated with the cytoplasmic face of the plasma membrane, but the examples of the lamins, and of the yeast mating factors (see table), are a reminder that proteins in various cellular compartments, or even secreted to the environment, can be modified in this way. So how is selectivity in targeting achieved? One possibility is that there is a palmitate or myristate receptor in the appropriate site. Another is suggested by the example of the *ras* proteins: the CAAX sequence could require cleavage carboxy-terminal to the cysteine to be acylated, as noted earlier. This is not always the case, as the primary structure of the related yeast YPT1 protein, which is involved in the secretory pathway, contains two cysteine residues at the carboxy terminus, either of which can allow acylation. The apparent three-residue extension seen in *ras* may thus be a 'pro' sequence which allows transport of the protein to the appropriate location for proteolytic activation of its acylation substrate potential. Such a two-step modification mechanism would ensure that inappropriate translocations would be avoided.

Although speculation concerning the roles of acylation tend to focus on membrane tethering, note that the secretory pathway has a requirement for palmitoyl-CoA¹¹ and that several viral fusion proteins are known to be acylated. Perhaps acylation is directly involved in membrane budding and fusion. In this case, one of the keys to understanding the dynamics of these processes could be the enzymatic control of acyl group addition and removal, as well as the identification of modified substrates. □

1. Low, M.G. & Saltiel, A.R. *Science* **239**, 268–275 (1988).
2. Sefton, B.M. & Buss, J.E. *J. Cell Biol.* **104**, 1449–1453 (1987).
3. Kaplan, J.M., Mardon, G., Bishop, J.M. & Varmus, H.E. *Molec. cell. Biol.* **8**, 2435–2441 (1988).
4. Towler, D.A. *et al. J. biol. Chem.* **263**, 1784–1790 (1988).
5. Aderem, A.A. *et al. Nature* **332**, 362–364 (1988).
6. Tamanoi, F. *et al. J. cell. Biochem.* **36**, 261–273 (1988).
7. Powers, S. *et al. Cell* **47**, 413–422 (1986).
8. Clarke, S., Vogel, J.P., Deschenes, R.J. & Stock, J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4643–4647 (1988).
9. Magee, A.L., Gutierrez, L., McKay, I.A., Marshall, C.J. & Hall, A. *EMBO J.* **6**, 3353–3357 (1987).
10. Wolda, S.L. & Glomset, J.A. *J. biol. Chem.* **263**, 5997–6000 (1988).
11. Glick, B.S. & Rothman, J.E. *Nature* **326**, 309–312 (1987).

Tony Magee is at the National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, and Michael Hanley is at the MRC Molecular Neurobiology Unit, Medical Research Council Centre, University Medical School, Cambridge CB2 2QH, UK.

Semiconductor devices

Conducting-polymer electronics

David Bloor

POLYMERS with electrical conductivities as good as those of metals are well known, but have yet to replace conventional electronic components because of difficulties in preparing and manipulating them. The development of soluble precursor polymers which can be purified and manipulated before being converted into a conductive form has invigorated the field. By using polyacetylene prepared by such a route, Burroughes *et al.* report on page 137 of this issue¹ the fabrication of polymeric semiconductor devices, transistors and diodes with characteristics superior to any previously reported. These devices operate via mechanisms intrinsically different from those in conventional semiconductor devices. Although the results supported by Burroughes *et al.* do not yet pose a threat to established semiconductor technology, they offer the prospect of novel devices which cannot be constructed with conventional semiconductors.

Polyacetylene is both the chemically

simplest and the most thoroughly investigated conductive polymer. The conjugated polymer chain comprises a skeleton of σ -bonded carbon atoms with a set of carbon p-electrons filling orbitals orthogonal to the plane of the chain. These form π -orbitals giving rise to an alternation of double and single bonds along the polymer chain. The π -electrons are localized in the double bonds but can be excited into delocalized π -electron states. Thus polyacetylene is a semiconductor with an energy gap of about 1.5 eV (electron volts) between the valence and conduction bands. This gap is too large for a significant number of electrons to be excited thermally into the conductive states at room temperature. Thus pure polyacetylene is only poorly conducting. The addition of strong electron accepting or donating species leads to a dramatic increase in conductivity. But Burroughes *et al.* use a very different means of making their samples conducting. ▶

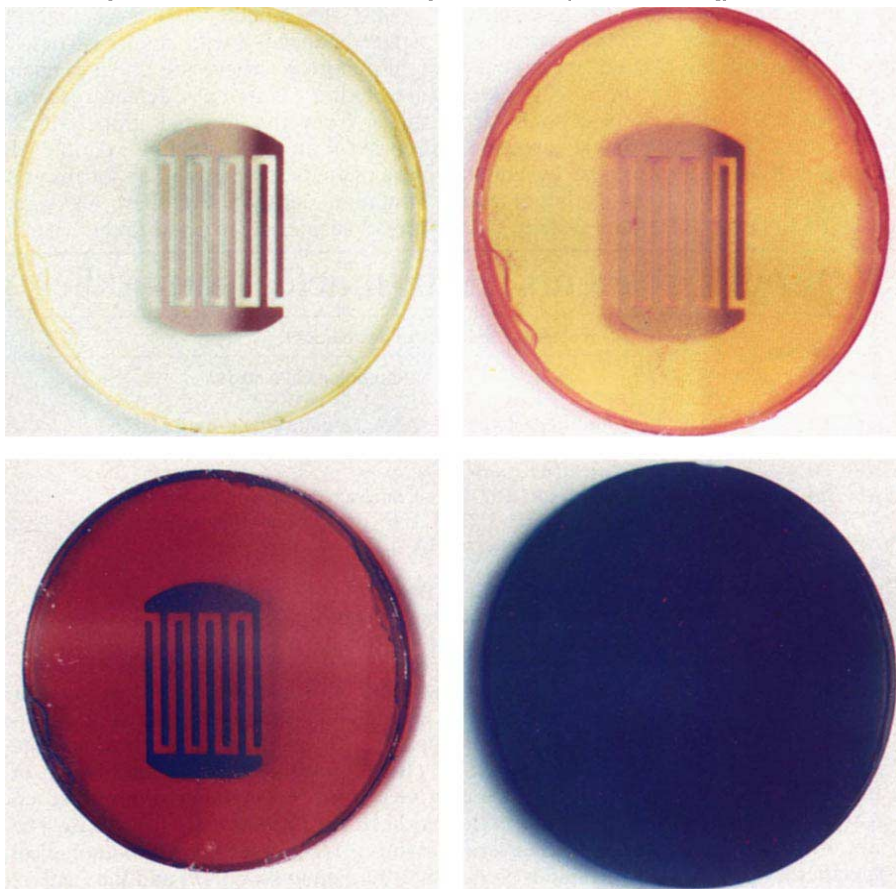


Fig. 1 Various stages in the thermal conversion of the Durham precursor polymer to polyacetylene (see also Fig. 4). The precursor is not conjugated and therefore colourless. As it is heated, sidegroups are eliminated and regions of conjugated 'polyacetylene' appear on the chain. These show a colour progression of yellow through red to black as the chain length increases. The polymer film here has been deposited on a substrate of silica, with gold interdigitated source and drain contacts pre-evaporated. (Courtesy of R. H. Friend.)

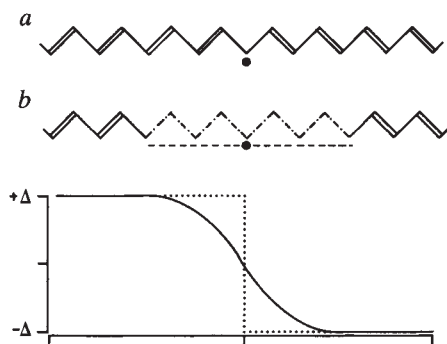


Fig. 2 *Trans* polyacetylene chain with a localized (Pöple-Walmsley²) bond-alternation defect (a) and a delocalized (Su, Schrieffer and Heeger³) soliton (b). The change in phase (Δ) of the bond alternation for the two species is shown at the bottom of the figure.

Despite the apparent analogy with inorganic semiconductors such as silicon and germanium — indeed, the oxidation and reduction of polyacetylene is termed 'doping' — there are fundamental differences. The atoms in inorganic semiconductors are fixed in a three-dimensional lattice of strong chemical bonds. Thus the deformation produced by an injected electron or near an electron and hole (empty electron state in the valence band), produced by photo-excitation, is very small. In consequence the band structure calculated in the absence of excitation can be used to describe the excited states — the rigid-band model. In an ideal polyacetylene chain the chemical bonds are weaker and lie along a single direction. Thus the deformation caused by added and excited charge is much larger and gives rise to new states of the system.

In the case of an ideal polyacetylene chain, the alternation of double and single bonds can be reversed without changing the total energy. But the π -electron at a sharp boundary where the bond alternation is reversed, would produce a paramagnetic defect (Fig. 2). This was recognized theoretically long before its practical significance was realized². Su, Schrieffer and Heeger³ developed the theory and showed that the change in phase of the bond alternation need not be sharp. This extended spin defect can move along the ideal chain with a very small activation energy. It also adds an energy level between the valence and conduction bands of the polymer, which produces sub-band-gap optical absorption and has unusual spin-charge relationships. The electrically neutral defect has an unpaired electron with spin 1/2, but charged states with zero spin that can conduct are produced by either removing or adding an electron (Fig. 3). Because the hamiltonian equation introduced by Su, Schrieffer and Heeger to describe the defect is analogous to other nonlinear wave equations, the defect is termed a soliton.

A lively debate has evolved concerning the existence of solitons^{4,5}. Experiments provide evidence both in favour of and against the model. It is clear that the defects associated with charge carriers in real materials, where the polymer chains are neither isolated nor defect-free, must differ from the original model of Su, Schrieffer and Heeger. For example, the paramagnetic defects will be localized by structural defects⁶. But there are signatures, particularly in photo-induced infra-

red spectra⁷, that indicate that the structural relaxation associated with charge excitations in polyacetylene is similar to that in the ideal soliton.

The presence of soliton mid-gap states should influence the behaviour of semiconductor devices constructed with polyacetylene. The device performance realized with polymer prepared directly by catalysed polymerization of acetylene has been too poor for this to be demonstrated⁸. Control over purity and morphology is provided by the 'Durham' precursor route^{9,10} (Figs 1 and 4). The soluble precursor polymer is purified by repeated precipitation and deposited by spin coating to give large-area films of uniform thickness. Burroughes *et al.*¹ use this approach to produce polyacetylene devices — diodes and transistors — with

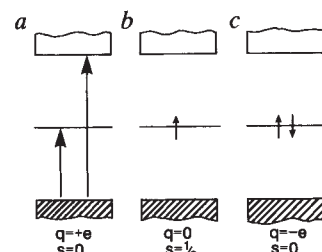


Fig. 3 Charge-spin relationships for different occupancy of the mid-gap energy level of a solitonic defect on a polyacetylene chain. Transitions to and from the mid-gap state give rise to sub-band-gap absorption, illustrated for the positively charged soliton at left.

performances several orders of magnitude better than those reported by previous workers.

They demonstrate that charge depletion and accumulation occurs at polymer-metal and polymer-insulator interfaces under the influence of an applied voltage in a way analogous with conventional silicon semiconductor-metal and -insulator interfaces. Thus in the metal-insulator-semiconductor field-effect transistor (MISFET; see Fig. 1 of Burroughes *et al.* on page 138), a voltage applied across the insulator layer causes charge to accumulate at or to be removed from the polymer-insulator interface. This changes the occupancy of the mid-gap levels, rendering the polymer conducting (Fig. 3), so that a current can pass between the source and drain electrodes attached to it. The change in occupancy of the mid-gap states can be monitored by the changing infrared absorption spectrum of the device as the fields are altered.

Thus the way of introducing charge

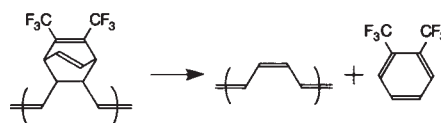


Fig. 4 Durham precursor route to polyacetylene, conversion is effected thermally and also results in the transformation of the initial *cis* polymer into the *trans* form (see Fig. 1).

Reference reagents for antinuclear antibodies

Reference antinuclear antibodies	
Reagent	Immunological features
AF/CDC1	Homogeneous pattern ANA. Doubles as anti-native DNA
AF/CDC2	Anti-SS-B/La
AF/CDC3	Speckled pattern ANA
AF/CDC4	Anti-U1 RNP (U1 small nuclear ribonucleoprotein)
AF/CDC5	Anti-Sm (U1, U2, U5, U4/6 small nuclear ribonucleoproteins)
AF/CDC6	Nucleolar pattern ANA
AF/CDC7	Anti-SS-A/Ro
AF/CDC8	Centromere pattern ANA
AF/CDC9	Anti-Scl-70 (DNA topoisomerase I)
AF/CDC10	Anti-Jo-1 (histidyl-tRNA synthetase)

AUTOANTIBODIES to nuclear and other intracellular antigens, often called antinuclear antibodies (ANAs), are useful as diagnostic aids in clinical medicine and as immunological probes in molecular and cell biology. Two international organizations, the International League Against Rheumatism (ILAR) and the International Union of Immunological Societies (IUIS), have now set up an International Committee on Antinuclear Antibodies at a meeting at the World Health Organization in Geneva. The committee will standardize reference sera containing ANAs of different immunological specificities and, in collaboration with the Arthritis Foundation (AF) and the Centers for Disease Control (CDC), will make these reference reagents available to clinical laboratories and research investigators. The table shows the reference ANAs now available. They can be obtained from the AF/CDC ANA Reference Laboratory, Immunology Branch, 1-1202 A25, Centers for Disease Control, Atlanta, Georgia 30333, USA.



100 years ago

A CORRESPONDENT of the *Daily News* in Lucerne sends to that paper an account of an electric mountain railway — the first of its kind — which has recently been opened to the public at the Burgenstock, near Lucerne. Hitherto it has been considered impossible to construct a funicular mountain railway with a curve; but the new line up the Burgenstock has achieved that feat under the superintendence of Mr. Abt, the Swiss electrical engineer. The rails describe one grand curve formed upon an angle of 112° , and the journey is made as steadily and smoothly as upon any of the straight funiculars previously constructed. A bed has been cut, for the most part out of the solid rock, in the mountain-side from the shore of the Lake of Lucerne to the height of the Burgenstock — 1330 feet above its level, and 2860 feet above the level of the sea. Through the opposition of the Swiss Government, each car is at the present time only allowed to run the half distance, and they insist upon the passengers changing, in order, as they say, to avoid collision or accident. A number of journeys were made up and down the mountain in company with an engineer, and the experience is sufficient to prove that the prohibition is altogether unnecessary. This interesting undertaking has been carried out at a cost of £25,000.

From *Nature* 38, 453; 6 September 1888.

carriers into the polyacetylene in these devices is quite different from the usual methods of chemical or electrochemical doping. Furthermore, the underlying physics is intrinsically different from that of conventional semiconductor devices with the injected charge stored in mid-gap rather than band states. Another notable point is that the good device performance is achieved with an essentially amorphous material. This is possible because in polyacetylene the fixed chain defects weaken π -electron interactions, giving a larger energy gap. Thus, in contrast to conventional semiconductors, structural defects do not produce localized states in the energy gap that would degrade device performance. □

1. Burroughes, J.H., Jones, C.A. & Friend, R.H. *Nature* 335, 137–141 (1988).
2. Pople, J.A. & Walmsley, S.H. *Molec. Phys.* 5, 15–20 (1962).
3. Su, W.P., Schrieffer, J.R. & Heeger, A.J. *Phys. Rev. B* 22, 2099–2111 (1980).
4. Roth, S. J. *Phys., Paris* 44, (C3) 69–76 (1983).
5. Heeger, A.J. in *Handbook of Conducting Polymers* (ed. Shothin, T.J.) 729–756 (Marcel Dekker, New York, 1986).
6. Thomann, H., Jin, H. & Baker, G.L. *Phys. Rev. Lett.* 59, 509–512 (1987).
7. Friend, R.H., Schaffer, H.E., Heeger, A.J. & Bott, D.C. *J. Phys. C* 20, 6013–6023 (1987).
8. Kanicki, J. in *Handbook of Conducting Polymers* (ed. Shothin, T.J.) 543–659 (Marcel Dekker, New York, 1986).
9. Edward, J.H., Feast, J.W. & Bott, D.C. *Polymer* 25, 395–398 (1984).
10. Bott, D.C., Brown, C.S., Winter, J.N. & Barker, J. *Polymer* 28, 601–616 (1987).

David Bloor is in the Department of Physics, Queen Mary College, London E1 4NS, UK.

Radioastronomy

Methanol masers map new stars

R.J. Cohen

INTERSTELLAR methanol was first detected nearly two decades ago. Methanol radiation has been observed at over 100 different frequencies, including several masers (powerful radio emissions). But there was particular excitement over the discovery in 1987 of the 12-GHz (gigahertz) maser transition¹. Its brightness is exceeded by only the water 22-GHz masers and it is widespread in star-forming regions throughout our Galaxy. On page 149 of this issue², R.P. Norris and co-workers present the first radio maps of seven methanol masers. There is an unaccustomed urgency about these and other observations of the 12-GHz maser. It lies in a frequency band where there is no official protection for radioastronomy and satellite broadcasts are already disrupting observations in some areas.

Cosmic masers are excited gas clouds whose spectral-line radiation is greatly boosted by the process of stimulated emission. They are valuable probes of star-forming regions because their high intensities make them readily detectable. Their powerful radio beams pass essentially unhindered through the screen of interstellar gas and dust which obscures the young stars at visible wavelengths.

Radio interferometers can map the maser clouds with great precision (see the recent News and Views article by Peter Scheuer³). The previously known hydroxyl (OH) and water masers have been extensively mapped using the Very Large Array (VLA) and MERLIN interferometers and very-long-baseline interferometry (VLBI) on angular scales from 1 arc-second down to 1 milliarcsecond. The picture revealed is fascinating but incomplete. Now radio astronomers are hoping that the new methanol masers will help to fill some of the gaps.

The maps of seven source regions presented by Norris *et al.* in this issue² give the first picture of the spatial distribution of methanol masers. Each region consists of a cluster or string of point-like masers spread over some 10^{14} m — the size of the Solar System — surrounding a massive young star. The methanol clusters do not look exactly like the hydroxyl or water maser clusters, so that they will reveal something new about star-forming regions, although it is too early yet to predict just what their contribution will be. There is a broad similarity with the hydroxyl masers, both in terms of the size of the maser region and the range of radial velocities. But the lack of detailed agreement suggests that there are physical differences, at least on the small scale.

Norris *et al.* remark on the tendency for the methanol masers to lie along lines, which could indicate an association with shock fronts, as has also been suggested for hydroxyl masers⁴. In a paper submitted to the *Astrophysical Journal* just 13 days after Norris *et al.* submitted theirs, Menten *et al.*⁵ present the first VLBI maps of 12-GHz methanol masers which have higher resolution. Their measurements of the absolute positions confirm the close association between methanol and hydroxyl masers and compact regions of ionized hydrogen (H II).

With their smaller beam, Menten *et al.* can establish that individual methanol masers have apparent sizes of less than 3×10^{11} m and brightness temperatures exceeding 2×10^{10} K. This measure of the maser gain is important for elucidating maser physics. There are strong indications that far-infrared radiation from the dusty cocoon of the young star is an important factor. For example Kembell *et al.*⁶ have found methanol masers in a survey of IRAS (Infrared Astronomical Satellite) far-infrared sources and they point out that sufficient far-infrared photons are available to excite the methanol masers in all cases. But other factors, perhaps chemical, must also be involved. As yet methanol masers have been found only in star-forming regions, not for example in cool circumstellar envelopes where similar infrared fields prevail⁷. Given the complexity of methanol's rotational and torsional spectrum it may be some time before the maser pump is fully understood.

The rapid progress which is being made in the study of the methanol 12-GHz masers is just as well. The 12-GHz frequency is marked out for satellite broadcasting. Although methanol masers are intrinsically very powerful, the signals which reach us stand no chance against man-made transmissions. This hazard was already apparent when the discovery of the 12-GHz maser was announced¹. Now radio astronomers are working against the clock to learn the secrets of methanol masers before progress of another kind drowns their cosmic message in a sea of man-made radio waves. □

1. Batrla, W. *et al. Nature* 326, 49–51 (1987).
2. Norris, R.P. *et al. Nature* 335, 149–151 (1988).
3. Scheuer, P. *Nature* 334, 107 (1988).
4. Elitzur, M. & de Jong, T. *Astr. Astrophys.* 67, 323 (1978).
5. Menten, K.M. *et al. Astrophys. J.* (in the press).
6. Kembell, A.J., Gaylard, M.J. & Nicolson, G.D. *Astrophys. J.* (in the press).
7. Norris, R.P. *et al. Astrophys. J.* 321, L159–L162 (1987).

R.J. Cohen is at the Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield SK11 9DL, UK.

Evolution

Perils of molecular introspection

Joe Felsenstein

THE rising flood of nucleic-acid sequence data has reopened old controversies on the evolutionary affinities of organisms, controversies that once seemed unresolvable. Sequences are being collected to resolve century-old tangles in the relationships of groups such as the lower invertebrates, the angiosperms and the protists, and even to see back 3,000 million years to the interconnections of the oldest bacterial groups. The phylogeny of the apes has generated the most controversy of all. Holmquist and colleagues^{1,2} now back up the claim, discussed by Diamond in a recent News and Views article³, that chimpanzees are more closely related to humans than to gorillas. They do this by analysing three data sets totalling 10,393 bases — sufficient to raise the question of whether new methods of analysis, such as Lake's method⁴ of "evolutionary parsimony", could resolve the issue.

Two computational methods have dominated the reconstruction of molecular phylogenies: parsimony and distance. The parsimony method finds the evolutionary tree that requires the fewest changes of nucleotides to explain evolution of the observed sequences. Distance methods compute a table of pairwise numbers of differences between sequences and try to fit this to expected pairwise distances computed from the tree.

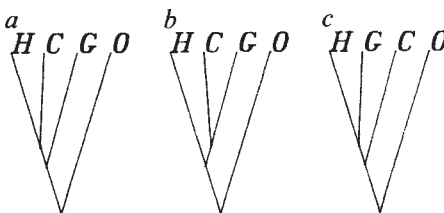
Both of these methods look at only part of the information in the data. For example, parsimony considers sites uninformative unless a different number of nucleotide changes are required on different trees. To use the full information present, maximum-likelihood methods have been developed⁵. These, however, are computationally difficult and are tied to a precise probability model of sequence change.

Lake's method is intended to cut the Gordian knot. It is simple to carry out and aims to avoid the problems which have plagued the other approaches. His application of it⁶ to resolving the affinities of the Archaeobacteria has been reviewed in the News and Views article by Penny⁷. Holmquist *et al.*^{1,2} use it to attack the difficult problem of ape phylogeny and find that humans and chimpanzees are the most closely related, the data having only a 3 per cent probability of showing a pattern this strong by chance.

Although Lake calls his method evolutionary parsimony, it is more closely related to likelihood methods. A similar method was independently developed by Cavender⁸, who coined the more precise term invariants. Invariant methods start

with nucleotide sequences for four species. Assuming that sites are evolving independently, they classify each site in the sequence into one of 256 possible patterns. In alphabetical order, these would be AAAA, AAAC, . . . TTTT where the four letters are the bases seen in the four species.

Any probability model of independent evolution at different sites predicts frequencies for these 256 classes. The form of the expressions depends on the shape of the true tree and the values depend on



The three phylogenetic trees that are in contention for the African apes (human, H, chimpanzee, C and gorilla, G). Most interest centres on a and b. The orang-utan, O, is included for comparison.

parameters including the branch lengths in the tree. Maximum likelihood methods are in effect ones which try to fit these 256 frequencies as closely as possible. Invariants try instead to detect certain of the regularities in the expected pattern frequencies.

Lake's invariants are sums and differences of expected frequencies of some of these patterns, expressions which will each be zero under two of the trees in the figure and not under the third. They achieve the status of invariance by being zero whatever the branch lengths in the tree.

Holmquist *et al.* do this tabulation for three nucleotide sequences for orang-utan, gorilla, chimpanzee and human. Of two classes of sites that are expected to have equal frequency of occurrence under the trees b and c in the figure, they find six sites in one class and none in the other. The occurrence of all six sites in one class is an event that has only 1 chance in 64 of occurring if trees b or c are correct, so that these alternatives to the human-chimp tree can be rejected.

Conventional parsimony would pay attention to a somewhat different collection of sites, the 'phylogenetically informative' sites that have only two nucleotides, each occurring twice. These fall naturally into three groups that each support one of the three trees. There are found to be 25, 13 and 16 of these, respectively. This again favours the human-chimp tree but less strongly than

do the invariants. Various statistical tests based on conventional parsimony^{9,11} fail to find statistical significance in this part of the data.

One advantage of Lake's invariants is that they are not misled by unequal rates of change at different sites and by unequal amounts of evolution in different branches of the tree. Their computational simplicity also makes them attractive to researchers weary of computers. But they too have their restrictive assumptions. It is critical to Lake's algebra that when a site which is now A undergoes a transversion, it must be equally likely to end up as a C or a T. It is not clear whether modest violations of this will be difficult for Lake's method.

Terms such as 'phylogenetically informative' mislead us by implying that all the relevant information is in a few sites. In fact, all sites contribute information. For example, if the human and chimpanzee are related and equally diverged from their ancestor, we would expect more sites with patterns like AACA than with patterns like ACAA, which is what is found. Both Lake's invariants and parsimony ignore this information, but likelihood and distance approaches do not.

We can either use all the information with a highly specific evolutionary model, as likelihood methods do, or trade some of that information for robustness by looking at a smaller subset of the data, as invariants, parsimony and distance methods each does in different ways. The addition of invariants to the phylogenetic arsenal at least seems to be persuading molecular evolutionists to take a broader look at the assumptions of their methods.

Even with the new analysis provided by Holmquist *et al.*^{1,2}, the ape issue is still far from being resolved. It is becoming clear that the phylogenetic tree of the African apes (human, chimpanzee and gorilla) is nearly a three-way split. Few physical anthropologists will be bowled over by a single test that merely reaches the 3 per cent level of significance. Nevertheless, that is the best yet done with sequence data. We are perhaps trying to squeeze blood out of stones but at least the stones are getting more numerous. Holmquist *et al.* show that they can be made somewhat moister as well. □

1. Holmquist, R., Miyamoto, M.M. & Goodman, M. *Molec. biol. Evol.* 5, 201–216 (1988).
2. Holmquist, R., Miyamoto, M.M. & Goodman, M. *Molec. biol. Evol.* 5, 217–236 (1988).
3. Diamond, J.M. *Nature* 332, 685–686 (1988).
4. Lake, J.A. *Molec. biol. Evol.* 4, 167–191 (1987).
5. Felsenstein, J. *J. molec. Evol.* 17, 368–376 (1981).
6. Lake, J.A. *Nature* 331, 184–186 (1988).
7. Penny, D. *Nature* 331, 111–112 (1988).
8. Cavender, J.A. & Felsenstein, J. *J. Classif.* 4, 57–71 (1987).
9. Felsenstein, J. *Syst. Zool.* 34, 152–161 (1985).
10. Felsenstein, J. *Evolution* 39, 783–791 (1985).
11. Templeton, A.R. *Evolution* 37, 221–224 (1983).

Joe Felsenstein is a professor in the Department of Genetics, University of Washington, Seattle, Washington 98195, USA.

Astronomy

Distant galaxy observed

R.F. Carswell

THE discovery of a radio galaxy at a redshift $z = 3.395$ is to be published by S. J. Lilly in the October issue of *Astrophysical Journal*. This is the most distant galaxy known and its properties give us some interesting insights into the nature of galaxies soon after they formed: its distance means that we are seeing the galaxy as it was billions of years ago. Only a few quasars with higher redshift ($z \approx 4$), representing more distant objects at an earlier epoch, are known (see the News and Views article by Peter Shaver, *Nature* **330**, 426; 1987). Of particular interest is the presence of a population of stars more than a billion (10^9) years old, suggesting that this galaxy and the stars in it formed very early in the history of the Universe.

The galaxy was selected for study because it is the source of strong radio emission, but it also seems to have some of the usual characteristics of a gas-rich galaxy. It shows short-wavelength continuum radiation indicating that a significant population of young stars is present, as well as line emission from the gaseous component and infrared continuum emission from an older stellar component. The continuum flux indicates that the total mass of stars is about 10^{12} solar masses and the angular size of the object leads to an inferred radius of the order of 10–25 kiloparsecs. Thus the density of material in this galaxy is quite similar to those observed nearby.

Perhaps the most important result of Lilly's observations is that the galaxy contains a significant number of stars likely to be more than a billion years old even when the light we see now was emitted. Thus the galaxy must have formed stars at an early stage in the evolution of the Universe. This is compatible with most models of galaxy formation, although it could constrain some variants of the biased-galaxy-formation picture (see Dekel, A. & Rees, M. J. *Nature* **326**, 455–462; 1987). In models that involve cold, dark matter, for example, galaxies themselves do not form until epochs corresponding to redshift $z \approx 4$ or later, though subgalactic masses could have formed earlier. (The redshift is the relative shift in light from a distant object due to the Doppler effect, which measures the recession velocity and, through the expansion of the Universe, gives an estimate of distance to the object.) If the initial burst of star formation occurs when the galaxy is formed, rather than earlier, it would be difficult to reconcile the presence of this galaxy with a relatively old stellar population with such a model except as a special case. If other

similar galaxies are found, models in which most of the galaxies form very early (at high redshifts, $z \geq 30$) from a spectrum of initial perturbations would gain support.

The presence of old stars at such an early stage can also be used to infer a useful lower limit to the age of the Universe. Present estimates of the expansion rate of the Universe give a Hubble constant somewhere around 50–100 km s⁻¹ per megaparsec (Mpc), with a preferred value near the middle of the range. Depending on the mass of the Universe, a Hubble constant of 75 km s⁻¹ Mpc⁻¹ corresponds to an age somewhere in the range $9\text{--}13 \times 10^9$ years. The lower value applies if the density in the Universe is critical, that is just enough for the expansion rate asymptotically to approach zero as the time approaches $t = \infty$, because under these circumstances the expansion rate must have been greater in the past. The higher value corresponds to a nearly empty Universe, where the deceleration due to gravity is negligible.

If the smaller value for the age of the Universe now is correct, then at redshift $z = 3.4$ the Universe was about a billion years old. Under these circumstances, it is hard to see how there could be any stellar component with an age of about a billion years or more. Of course, if the Universe is older (and so the expansion rate lower), or its density is lower, this time constraint is less of a problem. But it is difficult to reconcile this new discovery with an age of the Universe less than about 10 billion years even if galaxy and star formation were extremely rapid.

It is, of course, dangerous to draw sweeping conclusions from inferences based on difficult observations of a single faint object, but there is independent evidence suggesting that significant nuclear processing took place in stars at epochs corresponding to redshifts $z \geq 3.5$. Quasars are seen at higher redshifts with what seems to be a normal mix of heavy elements, suggesting that nucleosynthesis has occurred in stars rather than some exotic environment, and studies of absorbing clouds in the direction of these quasars also show that significant nuclear processing has occurred at early epochs with the material being returned to the interstellar medium. Taken together, these observations suggest that we should be considering galaxy and star formation on timescales of the order of hundreds of millions of years rather than billions. □

R.F. Carswell is in the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.

Daedalus

Shiver my timbers!

FORESTRY is a very inefficient business. Trees grow very slowly and make poor photosynthetic use of the sunlight they intercept; logging and transporting them is very labour-intensive. And after all this, most of the resulting wood is simply torn up into tiny pieces and reassembled into products like paper, laminates and particle boards. Why not use smaller and more convenient plants to start with?

The reason, of course, is that wood is uniquely strong. It has been evolved by large trees, to enable them to spread their leaves as widely and as loftily as possible, despite ferocious side-loading from the wind and the steady pull of their own weight. No smaller plant would have optimized cellulose and lignin into such a structurally splendid material.

So Daedalus is taking the logic a step further. He is breeding dwarf trees with the same structural excellence as their big cousins. They must continue to need the mighty strength of wood, despite being only a few centimetres high. So he is cultivating them in a powerful centrifuge.

Daedalus's centrifugal greenhouse allows seedlings of pine and oak and spruce to be grown under many gravities. A central 'sun' provides light for photosynthesis and gamma-radiation to encourage mutations. The rate of rotation is chosen to collapse most of the seedlings in any generation under their own centrifugal weight; the few tough survivors seed the next generation, which is challenged with higher gravity still. The ultimate product will be a micro-tree whose wood is still fully as strong as that of its larger ancestors.

A stand of Daedalus's microtrees will resemble a tenacious, bushy grass. The tiny trees will grow rapidly, covering the ground tightly with a photosynthetically efficient canopy a few centimetres above it. They will be readily harvested by mowing — though considerable power will be needed to sever their tough stems. Daedalus is already devising ways of processing the mowings, such as grinding under water to break up the plant structure and separating the little leaves (a useful source of protein) by differential flotation. Further grinding will reduce the wood-fibre feedstock to the size needed for paper, chipboard or any product made of aggregated particles.

A 'wood meadow' of the new trees will be cultivated like any other crop. Once established, it should form a stable and self-sustaining ecology. Herbivores will find the microtrees too tough to eat and digest; weeds will languish under its coherent canopy; regular mowing will keep it in equilibrium. Traditional, expensive forestry will survive only where really needed, to provide big timbers. David Jones

Radioactive protein-labelling techniques

SIR—We wish to alert the scientific community to the existence of a problem which, though widespread, has heretofore gone largely unreported. The use of ^{35}S -labelled amino acids (methionine, cysteine) to label proteins for further study is an almost universal practice in laboratories conducting molecular and cell biology. We assume that the precautions to be taken when using ^{35}S , a low-energy β -particle emitter, are generally recognized. Generally unappreciated, however, is the fact that solutions of ^{35}S -labelled amino acids release a volatile radioactive component which can pose a containment problem. It is released from every batch of ^{35}S -amino acids we have tested, regardless of the manufacturer or specific amino acid (though there is some suggestion that less is released from ^{35}S -cysteine than from ^{35}S -methionine).

When and where is this radioactive component most likely to be encountered? When a fresh vial of 8 mCi ^{35}S -amino acids is thawed, without a lid, in a large, closed container, approximately 1 μCi ^{35}S , as determined by standard air-sampling techniques, is released into the air. This may result from product breakdown occurring during the freezing process (which is known to accelerate physicochemical breakdown). Further volatile material is generated when ^{35}S -amino acids are added to cell culture medium and incubated at 37 °C, regardless of whether cells are present; thus this component is not generated metabolically. If it is indeed the result of a chemical/physical breakdown, the addition of stabilizers to the amino acids should decrease the amount of this component produced.

The volatile component is very soluble in water; thus the water present in incubators used for cell culture can become contaminated. A total of 300,000 c.p.m. (measured by liquid scintillation counting) were found in 500 ml water consequent to the use of 2.5 mCi ^{35}S -methionine in a single, 6-h incubation. Because this water is continually evaporating, recondensing and running down the inside of the incubator door, all the surfaces in these incubators—trays, side walls, door and even the outside of other dishes of cells—may become contaminated. The rubber gasket sealing the door and the metal fan which recirculates the air inside the incubators were found to be so highly contaminated that this is readily detectable by a hand-held G-M monitor. Filters used to wipe 10-cm² areas inside the incubator picked up several thousand c.p.m.

Although it is embarrassing that few researchers have picked up this problem directly, it is also disturbing that the manufacturers of ^{35}S -amino acids do not include more specific warning in their product-information sheets. All three

main producers are aware of the potential for such a problem, yet none, until very recently, has found time to investigate it.

While we await the manufacturers' recommendations, there are several precautions that are easily taken and which we find to be very effective in limiting contamination. First, vials of ^{35}S -amino acids should be thawed in a fume hood using a needle through the rubber septum to vent the vial. Alternatively, and perhaps more effectively, a syringe packed with charcoal (similar to those used for ^{125}I) attached to such a needle could be used.

Second, as activated charcoal readily absorbs the volatile radioactivity, we have placed a 16×16 inch pressed-charcoal filter of a honeycomb-type design on the top shelf of our incubator. This filter does not affect the CO₂ equilibrium in the incubator and, judging by the routinely low c.p.m. of air samples but increasing c.p.m. absorbed in the filter, it effectively

decreases contamination in the air. For less frequent use, activated charcoal in a tray or wrapped in several tissues to make a small bag will decrease contamination of the incubator.

Third, the water inside incubators should be changed on a regular basis, ideally after each labelling.

What is the volatile radioactive component? Two likely candidates have been named by the manufacturers: SO₂ or CH₃SH. At least two companies are working on its identity, and on solutions to the contamination problem. We expect a more detailed description of both the volatile radioactive component and the hazards it may pose on product-specification sheets in the very near future.

JILL MEISENHOLDER

TONY HUNTER

*Molecular Biology and Virology
Laboratory,
The Salk Institute,
10010 N. Torrey Pines Road,
La Jolla,
California 92037, USA*

Taxonomic instability continues to irritate

SIR—In his article on taxonomic instability, D. L. Hawksworth¹ may have aimed at the wrong target. We agree that name changes resulting from antiquarian research are usually unnecessary and that little would be lost if they were eliminated. But in our experience by far the most tiresome changes arise from what Hawksworth charitably calls "advancement of scientific knowledge", but which are often better described as mere changes of opinion as to where or whether, in an assembly of species, a generic distinction should be placed.

Systemics has been bedevilled by a failure to agree on the purpose of taxonomy. To Linnaeus it was simply to give a name to a species, just like naming an individual in a human family. With the acceptance of evolutionary change came the hope that taxonomy might reflect phylogeny. Few realized what a Pandora's box was being opened up. For the two functions could be acceptably combined only if everyone held identical views not only on the genealogy of plant and animal life, but also of the closeness of the relationship between species of a genus.

The conception of the genus uniquely displays the inevitable ambiguity: on the one hand, like family, class and higher taxa, it is a noun without objective reality. It is only an abstract view, based on an opinion, of which characters are of greater or lesser importance in phylogeny and therefore of where taxonomic divisions should be placed. By contrast, the species is a noun representing a real entity, at least as far as sexually reproducing organisms are concerned. But, unlike all other higher taxa, the genus also forms part of

the name of the species. Thus, if a species is thought to be sufficiently misplaced among its fellows, its position cannot be remedied without changing its name, thereby messing up past literature, confusing the language with different nouns for the same entity and imposing additional burdens on the memory.

It is the generation of genera that requires suppression, even more urgently than the suppression of priority searchers. Whatever established taxonomists say, unless this ever-increasing problem can be dealt with, the volatility of generic names will continue unabated until every species is allocated to its own genus.

Fortunately the remedy is simple, but it will negate one of the current recommendations of taxonomic legislators. It is to promote a half-way house between the genus and the species, namely the despised subgenus, which, like all other taxa beyond the species, would not be part of the species name. This would enable species names, such as the present binomials, to remain immutable, as logically all nouns should be in scientific literature, while allowing taxonomists the freedom to move species into or out of whichever subgenera are available or which they think need to be created.

As an example of the usefulness of subgenera, one may cite Darwin's work² on six-plated barnacles comprising the genus *Balanus*. He realized, of course, that narrower groupings were possible and accordingly divided the genus into sections. Pilsbury named these sections as subgenera, yet the literature remained intact³. But Newman and Ross⁴, though adding little new to knowledge, renamed Pils-

bury's subgenera as genera, thereby creating unnecessary havoc in the literature and language. Had they created subgenera, as Pilsbury did, their work would have been valuable and unexceptionable.

Darwin once argued strongly against the rule of priority, claiming that the best description, not the first, should create the species name⁵. But he conceded defeat when it was put to him that the law of priority alone would stand as an objective criterion which all would accept, whereas a rule based on opinion would be unenforceable. Exactly the same argument can be used against the present freedom to invent new genera or to move species from one to another at will, as it can only be a matter of opinion as to which species need or need not be separated generically.

When Darwin was faced with new and unfamiliar names for his broad genus *Scalpellum* he rejected them⁶, commenting that such changes undermined the basis of classification — by which we believe he meant convenience to the body of naturalists. How inconvenient indeed has the system become when the current half life of many generic names is only about 30 years.

D. J. CRISP

School of Animal Biology,

G. E. FOGG

School of Ocean Sciences,
University College of North Wales,
Bangor, Gwynedd, LL57 2UW, UK

1. Harkness, D.L. *Nature* **334**, 301 (1988).
2. Darwin, C. *A monograph on the Cirripedia: The Balanidae, Verrucidae etc.* (R. Soc., London, 1854).
3. Pilsbury, H.A. *Bull.* 93 (Smithsonian, Washington, DC, 1916).
4. Newman, W.A. & Ross, A. *Revision of the Balanomorph Barnacles* (San Diego Soc. Nat. Hist., 1976).
5. Burkhart, F. & Smith, S. (eds) *The Correspondence of Charles Darwin*, Vol. 4. (Cambridge University Press, 1988).
6. Darwin, F. (ed.) *The Life and Letters of Charles Darwin* (Murray, London, 1877).

Molecular biology running into a cul-de-sac?

SIR—The answer to your question whether molecular biology is running into a dead end (*Nature* **335**, 11; 1988) is yes, if its purpose is to help understand the uniqueness of life and its functions. Your parallel with the quandaries of spectroscopy in the 1920s before their resolution by quantum mechanics is apt, you could have added the quandary of the Michelson–Morley effect: that was resolved by the advent of special relativity. In both cases revolutionary insights were needed, inspired by Mach's positivist principles.

Biology is faced by two similar quandaries, both of which question whether DNA is the exclusive bearer of heritable information or whether all heritable information is structurally engraved in macromolecules. The first arises from the disparity between molecular and organismal evolution, exemplified by the sibling species of *Tetrahymena*, which do not

interbreed, but are morphologically indistinguishable^{1,2}. Despite the similarity, there is remarkable variation between these sibling species at the molecular level and there seems to be almost a complete dissociation between the molecular and the morphological levels.

A complementary dissociation, but on a much shorter timescale, arises in differentiation. Somatic cells maintain their unique identities throughout a lifetime, although all (except lymphocytes) have the same DNA. The 'modern' view is that stable interactions arise between the genome and its microenvironment during development. But when the cells of a tissue are grown in monolayer culture, they sequentially lose differentiated functions³. Stable inheritance of the differentiated state therefore depends on the organized state of the tissue rather than the DNA and its immediate environment. The predictable behaviour of the whole — organism, tissue — and the unpredictability of the parts — the isolated cell — has an uncanny resemblance to the predictability of the ensemble in quantum mechanics and the unpredictability of individual atoms and subatomic particles.

The epigenetic dilemma of Lederberg⁴ questions the adequacy of molecular reductionism for the transmission of biological information. This was recognized by Niels Bohr, who believed that life must be considered an elementary fact that cannot be explained, just as the quantum of action, along with elementary particles, forms the foundation of atomic physics⁵. He also observed that "only by renouncing an explanation about life in the ordinary sense do we gain the possibility of accounting for its characteristics"⁶. Similarly, the geneticist Sewall Wright suggested we treat the whole cell as a single gene at a higher level of integration than the chromosomal genes⁷. But the DNA revolution led a generation of biologists to believe that the secret of life lay entirely in the structure and function of DNA.

This faith is misplaced and the reductionist programme must be supplemented with a new conceptual framework. This will develop if we learn to think and work at the level of the cell and higher, as Bohr and Wright suggested. An example of how this might be done builds from the observation that the multiplication rates of individual cells in a cell culture population are heterogeneous. The reason why the fastest growing clone does not take over the population lies at the cellular, rather than the molecular, level. The fastest growing clones tend to throw off slower growing subclones, and the slowest growing ones tend to throw off faster growing subclones⁸. So although we cannot predict the behaviour of any individual cell, the population as a whole maintains a constant growth rate. There is a resemblance to the

methodology of quantum mechanics here in substituting a probabilistic for a deterministic description.

A second example is the study of the neoplastic transformation of cells in culture which has recently focused on the activation of oncogenes, and on chromosomal rearrangements. Oncogenes are demonstrated in tumours by transfecting NIH 3T3 cells with tumour DNA and getting them to transform. But NIH 3T3 cultures undergo spontaneous transformation if left undisturbed long enough. This transformation occurs only under specified conditions, requires competent cells, and most — but not all — of recently transformed cells revert to the normal state when transferred⁹. This operational analysis of the spontaneous transformation shows it to be an epigenetic rather than a genetic process; that is, one highly dependent on specific environmental conditions and at first partly, but not fully, reversible at the cell population level¹⁰. This does not exclude the possibility that mutations in DNA occur at some stage during the development of a tumour, and even contribute to that development. But it emphasizes that the driving force is a process of continuous interaction between cell and environment which can be attacked only at that level.

Experimental results must be viewed in a new theoretical framework. Walter Elsasser proposes the existence of two tiers for maintenance and transmission of heritable information in organisms¹¹. The first is the familiar replication-readout of DNA, and is the province of reductionism. The second — the "holistic memory" — is related to epigenesis and accounts for the unique characteristics of life. He states that "causal chains cannot be traced beyond a terminal point because they are lost in the unfathomable complexity of the organism." Therefore, we are required to work with the living state, be it cell or organism. We must now develop a robust model system which will give epigenetic studies at the cell and population levels the convenience and quantitation that phage and bacterial systems provided for molecular genetics.

HARRY RUBIN

Department of Molecular Biology,
University of California,
Berkeley,
California 94720, USA

1. Nanney, D.L. *J. Protozool.* **24**, 27–35 (1977).
2. Nanney, D.L. *Bioscience* **32**, 783–788 (1982).
3. DePomerai, D.I. et al. *J. Embryol. exp. Morph.* **77**, 210–220 (1984).
4. Lederberg, J. in *Centenaire Pasteur* (Elsevier, New York, in the press).
5. Bohr, N. *Nature* **131**, 421–423; 457, 459 (1933).
6. Bohr, N. in *Atomic Physics and Human Knowledge*, 67–82 (Wiley, New York, 1958).
7. Wright, S. *Am. Nat.* **79**, 289–303 (1945).
8. Grundel, R. & Rubin, H. *J. Cell Sci.* (in the press).
9. Rubin, H. & Xu, K. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
10. Nanney, D.L. *Proc. natn. Acad. Sci. U.S.A.* **44**, 712 (1958).
11. Elsasser, W.M. *Reflections on a Theory of Organisms* (Orbis, Quebec, 1987).

Adrenal transplants for Huntington's disease?

SIR—More than 100 patients with Parkinson's disease have been transplanted with adrenal medulla tissue into the ventricular wall of the caudate nucleus¹⁻³. Although the results are controversial³, the original rationale was that the adrenal chromaffin cells when put into the brain would release dopamine in sufficient quantities to reverse the functional deficits produced by the disease, and was based on animal studies demonstrating some benefit using this technique in models of Parkinson's disease^{4,6}.

Recently, Allen and his colleagues have performed an adrenal medulla autograft, but this time in a patient with Huntington's disease (Vanderbilt University Medical Center press release, 3 March 1988; see ref. 7). The rationale for doing this operation was apparently their unpublished finding that adrenal medulla transplants into rat caudate nucleus could protect against excitotoxin-induced neurodegeneration⁷. Their hypothesis appears to be that the adrenal grafts would protect against excitotoxin-induced neurodegeneration in Huntington's patients.

Although recent theory suggests that the neurodegeneration of the striatum in Huntington's disease may be related to the production of endogenous excitotoxins⁸, there is no proof for such a hypothesis. If further experiments were to substantiate the role of an excitotoxin in the aetiology of Huntington's disease and to substantiate the neuroprotective effect of adrenal transplants, then the rationale for the procedure would appear sound. But there is no published study demonstrating any beneficial effect of adrenal medulla transplants in animal models of Huntington's disease. Allen and his colleagues have reported that neonatal striatal tissue can protect against the excitotoxin-induced lesions⁹, although we have not been able to observe such an effect¹⁰.

The use of transplantation techniques in the treatment of Huntington's disease is still in the investigative stage, requiring further extensive animal research. Indeed the adrenal transplant techniques could exacerbate a patient's condition. If the original rationale for using adrenal transplants in the treatment of Parkinson's disease is correct, namely the release of dopamine^{4,5}, then it is possible that adrenal transplants could exacerbate the symptoms of Huntington's disease^{11,12}. Furthermore, tissue transplanted into the caudate appears to disrupt the blood-brain barrier^{13,14} and the consequences of this effect in Huntington's disease is unknown. In addition, some of the neurosurgical techniques used for the Parkinson's patients, such as making a cavity into the caudate¹, could result in facilitating the progression of the disease by removing

some of the already degenerated caudate nucleus.

Huntington's disease is a severe neurodegenerative disorder that causes profound incapacity and eventual death^{12,13}. As in most terminal diseases, there is always the temptation to try some unproven but hopeful treatment. But is it ethical to perform a treatment that has hardly been tested in animal models, and where the data have not been assessed in peer-reviewed journals, let alone replicated by other laboratories? These questions are especially pertinent in this case as the rationale for the procedure suggests that the adrenal transplant would be expected only to halt or slow the progression of the disease, and may not reverse the neurodegeneration already present. Although research into a surgical treatment for the disease has not progressed to a level that merits clinical application, it is an area of great potential.

The studies investigating fetal brain tissue transplants in animal models of the disease suggest that they may be beneficial. The transplanted fetal striatal-ridge tissue survives, grows, integrates, and produces recovery of function in the animals^{16,17}. But many experiments need to be performed in rodents and non-human primates before clinical trials in humans are undertaken. Furthermore, the use of fetal tissue has its own ethical considerations^{3,17} and may not be a practical procedure for some time. Until animal experiments are performed by a number of institutions and have demonstrated that the procedure is likely to be of value, it is hoped that adrenal medulla autografting for Huntington's disease does not follow the same inappropriately accelerated course that is being seen with adrenal medulla transplants into the brains of patients with Parkinson's disease³.

PAUL R. SANBERG
ANDREW B. NORMAN

*Division of Neuroscience,
University of Cincinnati Medical Center,
Cincinnati,
Ohio 45267-0559, USA*

1. Madrazo, I. *et al.* *New Engl. J. Med.* **316**, 831 (1987).
2. Lindvall, O. *et al.* *Ann. Neurol.* **22**, 455–468 (1987).
3. Sladek, J.R. & Shoulson, I. *Science* **240**, 1386 (1988).
4. Freed, W.J. *Biol. Psychiat.* **18**, 1205–1267 (1983).
5. Freed, W.J. *Science* **241**, 275 (1988).
6. Sladek, J.R. & Gash, D.M. *J. Neurosurg.* **68**, 337 (1988).
7. *Science News* **133**, 268 (1988).
8. Sanberg, P.R. & Johnston, G.A. *Med. J. Austral.* **2**, 460–465 (1981).
9. Tulipan, N., Huang, S., Whetsell, W.O. & Allen, G.S. *Brain Res.* **377**, 163–167 (1986).
10. Nash, D.R., Kaplan, S.M., Norman, A.B. & Sanberg, P.R. *Soc. Neurosci. Abstr.* (in the press).
11. Sanberg, P.R. & Coyle, J.T. *CRC Crit. Rev. Clin. Neurobiol.* **1**, 1–44 (1981).
12. Martin, J.B. & Gusella, J.F. *New Engl. J. Med.* **315**, 1267–2732 (1986).
13. Norman, A.B. *et al.* *Prog. Brain Res.* (in the press).
14. Sanberg, P.R. *et al.* *Expl. Neurol.* (in the press).
15. Bjorklund, A. *et al.* *Trends Neurosci.* **10**, 509–516 (1988).
16. Sanberg, P.R., Calderon, S.F., Garver, D.L. & Norman, A.B. *Psychopharmac. Bull.* **23**, 476–482 (1988).
17. Mahowald, M.B. *et al.* *Science* **235**, 1307–1308 (1987).

Cystic fibrosis, fertility and birth intervals

SIR—The issue of heterozygote advantage associated with the recessive allele for cystic fibrosis (CF) has been discussed in Scientific Correspondence recently¹⁻⁵. Reports of high sex ratio in affected families⁶⁻⁸ and a negative relationship between sex ratio and parity⁸ suggest an effect on fertility, although claims for increased family size are rightly discounted on the grounds that recessive alleles are most frequently detected in large families⁹⁻¹¹. Any effect on reproduction must almost certainly operate through one sex only^{1,7,8}, but an apparent association between the mutant haplotype and the Y chromosome² has not been confirmed^{3,4}. Many alternative, or additional, explanations have been offered (see ref. 11).

The most careful analysis of fertility is that of Jorde and Lathrop¹¹, in which the families of grandparents of patients were compared with precisely matched controls. Uncorrected figures show an apparently increased family size associated with the CF allele, but this was negated by correction for ascertainment bias. Mean and median birth intervals were also not significantly different from those of controls, but comparison of values from carriers with 20 sets of controls shows that in 19 cases both the average interval between marriage and the first birth and the average interval between subsequent births was shorter in the carrier families. This proportion indicates a highly significant difference ($P < 1 \times 10^{-4}$) in the first and subsequent birth intervals.

Using the authors' own figures it can be calculated that in an average family of four children, in which one parent is a CF carrier, the CF allele is replicated more than 5 per cent faster than its normal homologue in a control family. Jorde and Lathrop ascribe this difference to the superior fertility of families selected on the basis of expression of a recessive allele, but did not check or elaborate on their assumption.

A carefully controlled comparison of birth intervals in carrier and normal families of similar size should be informative, as this would eliminate the ascertainment bias associated with family size.

D.J. PRITCHARD

*Department of Human Genetics,
The University of Newcastle upon Tyne,
Newcastle upon Tyne NE2 4AA, UK*

1. Pritchard, D.J. *Nature* **330**, 319 (1987).
2. Kitzis, A. *et al.* *Nature* **333**, 215 (1988).
3. Ten Kate, L.P. *et al.* *Nature* **334**, 20 (1988).
4. Tümmler, B. *et al.* *Nature* **334**, 110 (1988).
5. Gibson, J.P. *Nature* **334**, 20 (1988).
6. Danks, D.M. *et al.* *Ann. hum. Genet.* **28**, 323–356 (1965).
7. Gloria-Bottini, F. *et al.* *Hum. Genet.* **54**, 79–82 (1980).
8. Pritchard, D.J. *et al.* *Archs Dis. Childhood* **58**, 290 (1983).
9. Ten Kate, L.P. *Ann. Hum. Genet.* **40**, 287–297 (1977).
10. Wright, S.W. & Morton, N.E. *Am. J. hum. Genet.* **20**, 157–169 (1988).
11. Jorde, L.B. & Lathrop, G.H. *Am. J. hum. Genet.* **42**, 808–815 (1988).

When the world turned

Owen Gingerich

The Genesis of the Copernican World. By Hans Blumenberg. Translated by Robert M. Wallace. MIT Press: 1987. Pp. 772. \$40, £35.95.

A GREAT puzzle for copernican scholars and intellectual historians has been the timing of what is frequently called the 'scientific revolution'. Copernicus's heliocentric reform was not the result of some brilliant new observation nor the collapse of a decrepit, over-encumbered epicyclic astronomy. It was made in mind's eye, a new aesthetic vision of how to put the Universe together. As Galileo later put it, he couldn't admire enough those who accepted the heliocentric doctrine *despite* the evidence of their senses.

But if this is really true, why did the revolution come in 1543, and not a century or a millennium earlier? Some have looked towards external circumstances: the age of exploration, of printing, of ecclesiastical upheaval. Hans Blumenberg, in a 1975 work that has only now been translated from the German, argues that the important *kopernikanische Wende* — "Copernican turning" (not revolution) — was the result of a long, by no means obvious, period of philosophical preparation and reorientation.

The episode of Joshua commanding the Sun — not the Earth — to stand still was one of the scriptural passages often thrown up to confute Copernicus. But in the preceding decades, claims Blumenberg, it had just the opposite effect, helping prepare the way for Copernicus. Why? Because the time-honoured aristotelian philosophy had identified the continual flow of time with the eternal and constant time-keeping daily revolution of the heavens. For God to stop the diurnal motion for the battle of Gibeon would have been to stop time itself. Clearly, scriptural exegesis among the Scholastics required a new theory of time, but also, in Blumenberg's view, there could not have been a copernican turning without a change in that philosophy and its identity of time with the motion of the starry sphere.

This is the stuff that Blumenberg's book is made of. Some of it is provocative and stimulating. Most of it, however, is excessively turgid and often just plain dull. The reader is faced with sentences such as, "Phenomenology had been founded, around the turn of the century, against psychologism, and it culminated, in the last years of its founder Edmund Husserl (years in which he was prevented from exercising his influence), in a critique of physicalism" or "In the *Dialogo* Salviati had always appealed to anamnesis and

maieutics when fundamental questions of mechanics were at stake".

The central argument of this lengthy book, on the role of "the removal of the cosmic character of time", comes well past the half-way point. The scope of the analysis, however, made it mandatory for Blumenberg to ask earlier how Copernicus

ing that the copernican arrangement automatically places the planets from the Sun in order of these periods — surely the single most compelling aesthetic argument for heliocentrism — strikes me as missing the main point of the important cosmological Chapter 10 of *De revolutionibus*.

At the time of the copernican quinquacentennial in 1973, some of us were told that Blumenberg's previous, smaller volume, *Die kopernikanische Wende*, was having an impact in the German-speaking world similar to Thomas Kuhn's *The Structure of Scientific Revolutions* among the Anglophones. Clearly both authors have addressed a related issue, of para-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Wheels within wheels — the system of Copernicus.

arrived at his theory. His thesis is that Copernicus's starting point was the ambiguity of placement of Mercury and Venus in the ptolemaic system, that is, whether their epicycles should be always closer or always more distant than the Sun — or maybe, why not circling the Sun? It is certainly worth mentioning that Copernicus explicitly cites the system of Martianus Capella, in which Mercury and Venus do revolve around the Sun, but to build an entire structure on this point without carrying it further seems to me to be making a mountain out of a molehill. To say that for Ptolemy Mercury revolves around the Earth in 88 days, Venus in 225 and so on shows that Blumenberg was just not on top of the technical material. And to give the heliocentric periods without mention-

digm change, but with very different conclusions. Blumenberg mentions Kuhn only twice, once in connection with the idea of a paradigm, and once to express his doubts. Some years ago I looked for Blumenberg's book in Harvard's Widener Library, the world's largest university collection, but in vain. Are the intellectual historians of two great linguistic cultures passing each other like ships in the night? Now that Blumenberg's book is at last available in English, I can perhaps best summarize by saying that, in contrast, Kuhn's *Scientific Revolutions* is a paradigm of clarity. □

Owen Gingerich is a Professor in the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

Seeing stars

M.G. Edmunds

Astronomical Optics. By Daniel J. Schroeder. *Academic:1988. Pp.352. \$45, £28.50.*

EXCEPT for the bunsen burner, telescopes are probably the best-known of scientific apparatus. Their design has not fundamentally changed since Newton excited the Royal Society with his model in 1672. The main mirror size has increased by a factor of 200, but most large telescopes rely on the same principle of keeping two mirrors accurately spaced apart and steerable around the sky.

Instead of the spherical and flat optical surfaces used by Newton, modern telescopes use combinations of other conic (or almost-conic) sections. This allows minimization of the inevitable aberration, or blurring, of images of finite size, which was a possibility recognized by James Gregory before Newton built a working telescope but which could not be realized until mirror grinding and polishing had improved. The main part of this useful book is an introduction to the optical design of simple reflecting telescopes — in this context, most large telescope designs are simple optically — together with description of the design of associated instrumentation such as spectrographs and their cameras. The exposition is clear, with careful reference to the different sign conventions used by other authors.

Much optical design these days is done by using ray-tracing computer software packages. The analytical approach described by Schroeder is an essential tool for efficient and sensible use of such programs, although it would have been helpful if a few more illustrations of actual aberrations had been shown. The result of the author's involvement in design of the Hubble Space Telescope is a good description of application of the methods to an extremely highly specified optical system, the goal here being a wavefront error of only one-twentieth of a wavelength with a primary mirror of 2.4 m in size. Such high performance is demanded because of the lack of atmospheric blurring that the Space Telescope will enjoy. But space-based observation is necessarily very much more expensive than that from the ground, and the next generation of large optical telescopes of primary mirror size 8 m or more will, of financial necessity, be Earth-bound.

It is in this area that Schroeder's book could have been taken further. The manufacture, testing and support of large mirrors is the most difficult of the problems facing the builders of the new 'very large' telescopes, yet it is barely mentioned. There is an account of the

diffraction image from several primary mirrors carried on a single mount, but only one telescope, the Columbus 2×8 m, is likely to be built on this principle, and even the pioneering Multiple Mirror Telescope in Arizona is to have its six mirrors replaced by a single, monolithic mirror.

Also offering tantalizing, but as yet unfulfilled, promise is the idea of servo-controlled *adaptive* optics to correct the wavefront aberrations introduced by the atmosphere in real time. Although such methods have been applied in other optical systems, the faintness of astronomical sources introduces its own problems and for some time it may be possible to correct only in the infrared where tolerances are slacker. Nevertheless, improvements in image quality have profound implications for auxiliary instrument design and the ultimate performance of the telescopes. The use (and constraints) of optical fibres

as feeds from telescope foci to spectrographs, which can effectively increase the survey speed of a telescope by one or two orders of magnitude in allowing many objects to be observed simultaneously, is also not given the attention it deserves.

The topics that are covered in this book are covered well. As well as telescope design, the reader is gently guided through the veritable zoo of modern astronomical spectrograph types. This is a text which sets out the general design principles of astronomical optics in a palatable way, although additional knowledge of the limits on quality and size of available optical components will be required by those described in the foreword as the "future practitioners of the magic art of doing it with mirrors". □

M.G. Edmunds is in the Department of Physics, University of Wales College of Cardiff, Cardiff CF1 1XL, UK.

Fruits of industry

Norman Myers

Acidification in Tropical Countries, SCOPE 36. Edited by Henning Rodhe and Rafael Herrera. *Wiley:1988. Pp.405. \$155, £65.*

WE ALL know about so-called acid rain in Europe and North America. But what about the rest of the world? Is there the prospect elsewhere of widespread acidification of soils and water bodies, damage to vegetation, corrosion of buildings and the rest of the unfortunate syndrome? In particular, how far are tropical ecosystems becoming vulnerable to sulphur and nitrogen emissions?

Rodhe and Herrera, Swedish and Venezuelan scientists, respectively, have assembled an impressive array of contributions by experienced observers. As the editors stress, this is not a definitive work. Yet the early findings are sombre enough, as the book makes plain. Although the emission density of sulphur and nitrogen oxides in most tropical areas is still far less than in many temperate-zone industrial countries, several regions are likely to experience pronounced acidification if current trends in population growth, urbanization, industrialization and consumption of fossil or biomass energy continue beyond the turn of the century.

Much of the concern stems from the fact that tropical soils are often naturally acidic, low in cation-exchange capacity and high in aluminium concentration. This leaves them unusually susceptible to acidifying processes of human origin. The situation is exacerbated by a number of further characteristics of tropical environments. Daytime temperature and sunlight intensity are generally high throughout

the year, enhancing the efficiency of many photochemical processes in the atmosphere. Rainfall rates are high, at least seasonally, offering much scope for wet depositions of sulphur and nitrogen compounds. And many tropical winds, such as the trade winds, change little in direction over long periods of time, meaning that areas downwind of a source of air pollution can continuously be subject to high concentrations and deposition of pollutants. Rapid deforestation, now overtaking much of the humid tropics, is a further acidifying mechanism — forest cover is often replaced with bush and grass savannahs, and the subsequent fires, whether natural or anthropogenic, become another source of trace gases in the atmosphere.

The book starts with a consideration of which specific tropical areas are likely to be particularly sensitive to acidification because of their already acid soils. It then goes on to examine the regions in which emissions of sulphur and nitrogen pollutants are likely to increase. Six areas that are undergoing rapid industrialization, and/or have large urban communities with large numbers of motor vehicles are cited — southern China (in which additional acidification is already manifest), south-east Asia (Thailand, peninsular Malaysia, Sumatra and Java), south-western India, West Africa, south-eastern Brazil and northern Venezuela.

The findings described in the book are tentative, and will need to be elaborated with more data and analysis. But this is a timely assessment of an emergent problem that is likely to demand increasing attention from conservationists and environmental scientists. □

Norman Myers, Upper Meadow, Old Road, Headington, Oxford OX3 8SZ, UK, is a consultant on development and environmental matters.

Safe as houses? The risks of radon

John Fremlin

Radon and its Decay Products: Occurrence, Properties, and Health Effects. Edited by Philip K. Hopke. *American Chemical Society*: 1987. Pp.609. \$89.95, £63.85.

Indoor Radon and its Hazards. Edited by David Bodansky, Maurice A. Robkin and David R. Stadler. *University of Washington Press*: 1987. Pp.147. Hbk \$20, pbk \$9.95.

Environmental Radon. Edited by C. Richard Cothorn and James E. Smith, Jr. *Plenum*: 1987. Pp.363. \$55, £66.

Radon and its Decay Products in Indoor Air. Edited by William W. Nazaroff and Anthony V. Nero, Jr. *Wiley*: 1988. Pp.518. £65, \$75.

Health Risks of Radon and Other Internally Deposited Alpha-Emitters. BEIR IV. *National Academy Press*: 1988. Pp. 602. Hbk \$51, pbk \$39. In Europe distributed by Wiley, hbk £36.45, pbk £27.90.

THE general lack of public concern about the health risks caused by the natural radioactive gas radon is one of the odder features of modern life. Whatever the reasons for that indifference, little commensurate with the real importance of the subject has been published in the West. With the appearance of these five books, this curious and undesirable silence has now been broken. All are aimed at professionals who are familiar with the properties of radioactive materials and are interested in radiation hazards.

Each of the first four volumes consists of articles written by various experts. That edited by Hopke, for example, includes a useful overview by Hopke himself, and a series of updated papers from an American Chemical Society symposium held in April 1986. Unusually for a book with this provenance, the result is not only authoritative but readable and coherent.

BEIR IV differs in two main ways from the other volumes. First, the individuals who produced the text remain anonymous, but were members of a long-standing committee of 12 medical, radiological and epidemiological experts who were chosen "for their special competence and with regard to appropriate balance".

Secondly, it deals with other alpha emitters, ingested rather than inhaled, as well as with radon and its short-lived decay products. These other alpha emitters will usually be found in other organs than the bronchi and lungs and many of them have long half-lives. Most of the discussion of sub-cellular damage caused by alpha particles from plutonium, for example, applies also to the damage caused by alpha particles from the radon progeny. This receives scant attention in the other books apart from Hopke's.

The first practical requirement in looking at health hazards caused by radon is to determine the quantity of the progeny being inhaled. Several quick and effective methods, some of which need expensive equipment, are described in each of the first four books (in some detail in Hopke). Such methods are valuable for preliminary surveys, and for comparison of the hazards in different areas, but are inadequate for the estimation of the average risks to the population. The concentrations of radon progeny in any particular room on any particular day depend on whether doors and windows are open or shut; alter every time someone opens the door; vary between winter and

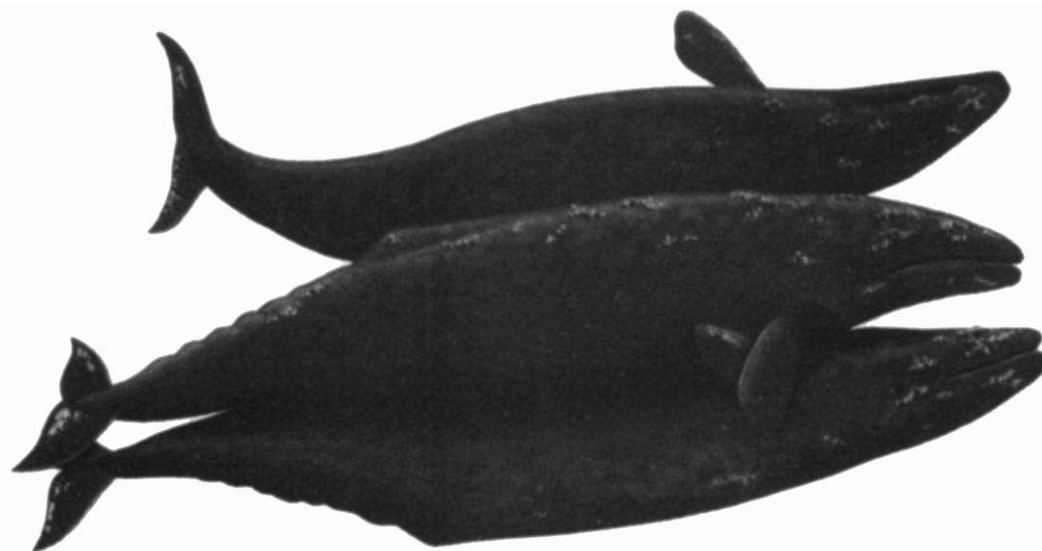
summer; and depend on the difference between indoor and outdoor temperature.

Hopke's book describes several national radon surveys employing different measurement techniques, including that used in the systematic British study in which C39 plastic alpha-track detectors were placed for two successive continuous six-month periods in 2,000 homes. From this, an average annual effective dose-equivalent of 43 millirems was deduced assuming 75 per cent mean occupancy.

A contributor to Cothorn and Smith describes a system in which six separate measurements are made at least four weeks apart, each of 100 hours duration, with care to ensure similar ventilation conditions. This was quite appropriate for following the decommissioning of uranium mill sites for which it was developed, but not for long continuous periods in large numbers of houses. The use of plastic alpha detectors, also covered briefly, is much more suitable.

The book of Bodansky *et al.* includes an account of such detectors used for 30 days at a time. These can of course be used one after another to cover a year or more, but the importance of obtaining measurements all round the year is not stressed. Thermoluminescent detectors are also mentioned; these have their advantages, but are less sensitive and add to their record the doses from cosmic rays and gamma rays from the ground. Nazaroff and Nero has a brief but adequate article on the etched-track method, together with descriptions of several other methods too expensive for use over long periods in large numbers of houses. BEIR IV does not cover the measurement of indoor radon at all.

Having assessed the doses received, the next step is to estimate the health risk. The method adopted by the International Commission for Radiological Protection (ICRP) is to use the numbers of excess



Menage à trois — mating for the grey whale Eschrichtius robustus is complex and elaborate. Only one male is involved in the actual mating; the other takes an upright position on the far side of the female, acting as a prop or wedge. The picture is taken from Lyall Watson's Whales of the World, just reissued by Hutchinson, price £11.95.

cancers caused by the atom bombs dropped on Japan and to assume that there is a linear dose-to-risk relation from colossal doses of 100 rems or more delivered in a fraction of a second to doses of a few rems delivered over decades. Bodansky *et al.* is the only book to consider this linearity hypothesis at length, and to point out the unlikelihood of linearity over such a range of dose and dose rate. However, the probable exaggeration of low-level risks gives me confidence that the legal limits based on the ICRP assumptions (100 millirems per year for the public) are not set too high.

In all the books, there is lengthy discussion of the alternative method of risk estimation by using the records of the excess cancer rates in uranium miners before the dangers of inhalation had been realized. The results are complicated because a large but unknown proportion of the miners were cigarette smokers. The cancer rate among white miners was much higher than that expected of cigarette smokers in general; but, as is explained in all the books except Cothorn and Smith's, which quotes conflicting reports, the cancer rates from smoking and from inhaling radon progeny are likely to have a synergistic interaction.

In Bodansky *et al.*, Hopke and BEIR IV, there is mention that a group of Navaho Indian miners, few of whom smoked, also suffered extra lung cancers. But the smaller numbers involved, and the uncertainty of the dose received, prevent any reliable estimates of the risk to non-smokers — except that there is a risk to non-smokers.

The results of the study of the uranium miners are less definite than might be expected. In the summary of his contribution to Hopke's book, Steinhäusler says:

Available input data for the risk assessment for low level radon daughter (Rn-d) exposure are mostly either of low quality, partially contradictory, or simply "guesstimates". Therefore at present only the upper limit of the risk can be estimated. . . . It is concluded that for "normal" indoor Rn-d exposure the resulting risk is negligible compared to other risks "accepted" by society.

In his longer and later article in Nazaroff and Nero, he expresses the same point in a different way: on the basis of the data on uranium miners, the most probable risk is ten per million person years per rem, $\pm 100\%$.

In estimating risks, there is also the question of differences of susceptibility among the populations exposed. Risks may be absolute (a given dose will add the same extra risk to everyone of a given age) or relative (the extra risk will be proportional to the pre-existing risk of lung cancer due to other causes). Data for average populations can be made to fit either hypothesis. A relative risk would imply

Grub's up

John Treherne

Why Not Eat Insects? By Vincent M. Holt. *British Museum (Natural History)/E.W. Classey: 1885/1988. Pp. 99. Pbk £3.95. In the United States available from ISBS, 5602 NE Hassalaw St, Portland, Oregon 97213, \$7.95.*

WHEN I was a child, a rather fierce relative used to boast that he ate locusts during the Matabele Rebellion (or it may have been the Boer War, I forget which). They tasted very nice, he said: like shrimps, when they were cooked properly.

Uncle Dennis had to eat the beastly things or he would have starved. And that is the odd thing about Mr Holt's little book, for there is no indication of *why* he actually started to eat insects. Did the other boys make him do it at school? Or did he survive on them in the Gobi Desert and then found it difficult to kick the habit? We are never told. But we are given cogent reasons why we should tuck into the Insecta. There are the poor, for example. A great boon for them in 1885, when Mr Holt originally penned his study, as well as, presumably, for their modern equivalents, teetering on the verge of malnutrition with junk food. As Mr Holt

that cigarette smokers, who have a ten times larger risk, in Europe and the United States, of contracting lung cancer than have non-smokers, will also suffer a ten times larger excess for the same dose from the radon progeny. This possibility is discussed in all of the books except that edited by Hopke. BEIR IV comes down in favour of the relative risk model, and is likely to be right, but one cannot yet be convinced of that.

All of the books, except BEIR IV, include descriptions of varying length of the means by which radon doses can be reduced. For new buildings, sealing the foundations, including the entry points of service pipes, or provision of a well-ventilated crawl space below the ground floor, will be effective and involve no subsequent costs. For existing buildings, too much depends on the type of structure, its heating system and the way of life of its inhabitants to give generally useful recommendations. Keeping the windows open all the year round and wearing extra clothes can be very effective if (as at my age, I do not) you prefer cold to radon.

All the books are commendably free of printing errors, but the references have the serious disadvantage of being listed chapter by chapter — an infuriating practice. As usual, a variety of units are used: rads and rems, grays and sieverts, and

points out, what a pleasant change from the labourer's unvarying diet — in his day, of bread and lard and bacon — would be a good dish of fried cockchafters or grasshoppers. And, in so varying the monotony of their diet, the poor would unwittingly assist agricultural pest control. No need to wash the vegetables either, for what could be nicer than cabbage served up with a delicately flavoured fringe of caterpillars?

But it is to gourmets that Mr Holt primarily appeals. After all, they have been guzzling the Invertebrata more than most — snails, lobsters, oysters and the like — and just might be tempted by stag beetle larvae on toast or new carrots with wire-worm sauce or gooseberry cream with sawflies. Yet Mr Holt's ideas seem to have made little impact on the society of Oscar Wilde and Mr Gladstone. They might fare better in Mrs Thatcher's Britain though. Not with *her* poor either, I suspect, but it is just conceivable that the jaded palate of the upwardly affluent might be tickled by *Larves de Guêpes Frites au Rayon* or *Boeuf aux Chenilles* — and could even lead to the setting up of the odd entomological brasserie in Kensington or Holland Park. □

John Treherne is a Fellow of Downing College, Cambridge CB2 1DQ, UK, and Honorary Director of the AFRC Unit of Insect Neurophysiology and Pharmacology, Department of Zoology, University of Cambridge.

working level months (WLMs). It is a great pity that the inconveniently large gray and sievert were ever invented, but the incommensurable WLMs met a real need. I have throughout translated all units into rems, counting 1 WLM as 1 rem; the error involved is small compared to the uncertainties in the data for which it was used.

The message of these five books is that a great deal of research has produced little evidence to support or to weaken the estimates of risk made by the ICRP. It has, however, confirmed that the radiation doses from radon exceed by a factor of many hundreds those arising from the nuclear power industry — hence my puzzlement over why the public bothers so little about the health risks of radon.

It will take a different type of publication to get that point through to the public. For professionals, each of the volumes reviewed here says something better than do the other four. For myself, I prefer that edited by Bodansky *et al.*, the only one to compare the dangers of radon and its progeny with other radiation hazards, and which presents its information in the most succinct and digestible form. □

John Fremlin, 46 Vernon Road, Edgbaston, Birmingham B16 9SH, UK, is Professor Emeritus of Applied Radioactivity in the University of Birmingham.

From continental drift to plate tectonics

Henry Frankel

Twenty-five years ago, on 7 September 1963, a paper by F.J. Vine and D.H. Matthews, "Magnetic Anomalies Over Oceanic Ridges", appeared in *Nature*. It attracted little immediate interest. But the paper turned out to be one of the seminal contributions to a sea-change in the Earth sciences that brewed through the 1950s and early 1960s and culminated in the theory of plate tectonics.

THE Earth sciences underwent a revolution in the 1960s which rocked the subject to its core. The revolution culminated in the birth of plate tectonics, and ended the 55-year-old controversy over continental drift. Central to events was the Vine-Matthews-Morley hypothesis. This hypothesis, independently invented by Fred Vine and Lawrence Morley, led to the confirmation of seafloor spreading, then the most modern form of continental drift. Before acceptance of the hypothesis few Earth scientists believed in continental drift; with its confirmation, few dared to argue against its modern forms.

Although some prominent figures such as Arthur Holmes, Lester King and S.W. Carey defended continental drift during the early 1950s, the vast majority of Earth scientists believed that it was a relic of past geological thinking. In this rather hostile climate, two groups of (primarily British) palaeomagnetists developed a new case for continental drift.

Palaeomagnetism

The first group, begun by S.K. Runcorn, was initially housed in the Department of Geodesy and Geophysics at the University of Cambridge, and the second, under P.M.S. Blackett, found its first home in the Physics Department at the University of Manchester. Key early members of Runcorn's group were J. Hospers, E. Irving and K. Creer. R.A. Fisher, the great mathematical biologist, provided them with the statistical package needed to analyse the scatter endemic to palaeomagnetic data. Runcorn and his colleagues began to attack the problem of developing a palaeomagnetic procedure for determining the positions of the continents throughout geological time, soon after Irving, the first among the British palaeomagnetists, realized how palaeomagnetism could be used to test drift theory. Runcorn, who had the foresight to include geologists as well as physicists in his team, hired Irving in June 1951 to help collect samples. When Irving

Here, Henry Frankel describes the genesis and influence of the work of Vine and the contributions of some of the many Earth scientists who participated in the transformation of their discipline. In the following Review Article (p. 131), Peter Molnar picks up the story and reviews the progress made in applying plate tectonics to the thorny question of continental deformation.

told Runcorn of the connection between continental drift and palaeomagnetism in the autumn of 1951, and suggested that India would be an ideal area to test, Runcorn immediately saw its importance.

Within a few years, Creer presented the first polar wander curve, plotting the movement of the British Isles throughout the geological past, and Runcorn and Irving soon added polar wander curves

"Do you know if *Nature* gets their articles reviewed, or do they publish anything?" . . . "Well, we're just about to find out because, you know, I just put my paper in, and if they publish that they'll publish anything".

from North America and Australia. Hospers and Runcorn also put forward arguments in support of one of the crucial assumptions in the palaeomagnetic case for drift, namely, the idea that the Earth's axis of rotation and its main magnetic field coincide. Meanwhile Clegg, hired by Blackett in late 1951 to begin palaeomagnetism at Manchester, found support for continental drift from his work in Britain. Later, Clegg and his co-workers extended Irving's Indian work. Finally, Irving, Runcorn and N. Opdyke, a new Runcorn recruit, and Blackett proposed palaeoclimatological arguments in corroboration of their palaeomagnetic work. Thus, by the late 1950s both groups were able to present a totally new defence of continental drift.

Palaeomagnetism rekindled interest in continental drift, but it changed few minds. Palaeomagnetists sometimes disagreed among themselves, and most non-palaeomagnetists ignored, viewed with suspicion or failed to understand the subject. Critics questioned its assumptions, and the reliability and completeness of its database. Nevertheless, palaeomagnetism influenced several of those who later played crucial roles in the revolution and showed, by example, that the combined efforts of geologists and geo-

physicists offered a promising way to attack problems in the Earth sciences.

Seafloor spreading

A stronger case for continental drift was needed. It came from marine geology, another relatively young field. By the early 1960s, the combined forces of several research teams had turned the sea floors from areas of deep mystery into places of backyard familiarity. The worldwide network of mid-ocean ridges was identified through the work of M. Ewing, B. Heezen and M. Hill. Although the data were often incomplete and variable, ridges were generally in the middle of ocean basins, had a central rift valley, and were characterized by shallow earthquakes, a central magnetic anomaly, somewhat anomalous seismic velocities, high heat flows and little sediment.

In the north-eastern Pacific, R.G. Mason and A.D. Raff uncovered a large and unexpected pattern of magnetic anomalies (Fig. 1, over). The pattern was made up of narrow, north-south-trending anomalies, exhibiting high magnetic readings, separated by similarly shaped and trending anomalies displaying low readings. Moreover, lengthy, east-west-trending fracture zones, previously discovered by H.W. Menard, divided the zebra pattern of magnetic anomalies into sections, and V. Vacquier noticed how sections bordering one of the fracture zones could be matched with little distortion, if it were supposed that sea floor had undergone lengthy horizontal displacement along the fracture zone (see Fig. 2). Investigation of sediments on other areas of the sea floor revealed two more surprises — their thickness turned out to be much less than had been expected, and, despite numerous dredge and core samples, researchers failed to find any older than Cretaceous (about 135 Myr old).

Special attention was directed towards the origin of ocean ridges. In retrospect, the explanation having the greatest influence was that proposed by Harry Hess of Princeton University. His 1960 solution, which he categorized as an essay in geopoetry, was labelled "sea-floor spreading" by R.S. Dietz in his own 1961 version of the idea. Hess proposed that ocean crust is created along ridge axes as magma is forced up from the mantle, and results, as he later argued, in ocean trenches along the periphery of ocean basins (Fig. 2). His hypothesis offered a qualitative solution to the origin of ocean ridges. Moreover, by restricting the life of the sea floor, he explained why no sediments older than Cretaceous had been recovered. ▶

Hess had not been a proponent of continental drift before advancing the concept of seafloor spreading. But he was not unhappy about grafting it onto drift theory. First, he had been impressed with the developments in palaeomagnetism. Secondly, he realized that his hypothesis offered 'drifters' a way around their chief difficulty, namely, how the continents could plough vast distances through the ocean floor. This 'mechanism' problem plagued almost every classical presentation of continental drift — the most notable exception, Arthur Holmes's 1931 version of drift, in many respects foreshadowed Hess's idea.

Vine–Matthews–Morley

The Vine–Matthews–Morley hypothesis is a corollary of seafloor spreading and reversals in the polarity of the Earth's magnetic field. The reasoning behind it is that as newly created seafloor material cools through its Curie point, it becomes permanently magnetized in the direction of the existing geomagnetic field, and thus, as the sea floor spreads, "blocks of alternately normal and reversely magnetized material would drift away from the centre of the ridge and parallel to the crest of it" (Fig. 3).

Vine came up with the hypothesis while Matthews, his supervisor, was away. Several factors prepared Vine for his inspiration. He was favourably inclined towards continental drift. His introduction to drift occurred when he was about 14, upon reading a geography textbook. The authors discussed the coastline fit between Africa and South America, remarking that geologists were undecided about its significance. Vine was intrigued. When an undergraduate geologist at Cambridge, he was wont to defend drift theory, and he was influenced by Brian Harland who favoured it. Because Vine had a strong background in geology, but was versed in mathematics and physics, he was sensitive to geological matters; and, unlike most geologists, he was able to understand and evaluate the palaeomagnetic arguments for drift and geolarity reversals. His mathematical abilities allowed him to understand a method for analysing magnetic anomaly data developed by K. Kunaratnam, a student of Mason's at Imperial College. Vine programmed these mathematics for the then relatively novel digital computer to analyse the data Matthews had collected from a survey of the Carlsberg Ridge.

There was, too, the influence of Matthews and Bullard at Cambridge. Both believed in continental drift and geolarity reversals, and encouraged

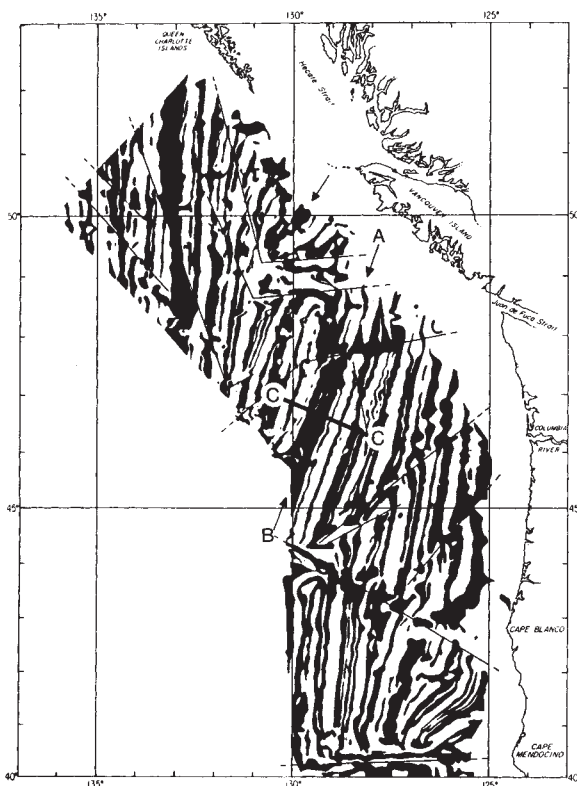


Fig. 1 Map of the north-eastern Pacific, showing the magnetic 'stripes'; A–B is the Juan de Fuca ridge (after Raff and Mason, 1961). Dark areas are regions of the sea floor that give anomalously high magnetic readings (positive anomalies); white areas are negative anomalies. In the Vine–Matthews–Morley hypothesis, these stripes are interpreted as regions of normally and reversely magnetized ocean crust (see Fig. 3).

Vine to make sense out of the marine magnetic data. Vine was also impressed with Hess and his ideas. Hess presented seafloor spreading to a student geological conference while visiting Cambridge in January 1962 when Vine was a third-year undergraduate. Impressed, Vine took notes and after Hess's visit addressed the Sedgwick Club, the geological club at Cambridge, on "HypotHESes". He argued that seafloor spreading was worthy of serious pursuit.

When it was proposed, the Vine–Matthews–Morley hypothesis met with indifference or scepticism. Morley's version was rejected by *Nature* and the *Journal of Geophysical Research*. In 1964 he tacked it onto a lengthy article on geochronology, which was published but went unnoticed. Even Vine and Matthews's article, which appeared in *Nature* on 7 September 1963, attracted little attention. Runcorn, for example, had not seen it, and so did not invite them to discuss their ideas at the 1964 Royal Society symposium on continental drift, which Blackett and Bullard asked him, the junior of the three, to co-organize.

The controversial nature of the hypothesis is indicated by the rejection of Morley's original paper and a comment Vine made before hearing about the acceptance of his paper. When asked "Do

you know if *Nature* gets their articles reviewed, or do they publish anything?", he said, "Well, we're just about to find out because, you know, I just put my paper in, and if they publish that they'll publish anything". He later remarked that when the hypothesis was proposed, "it created more problems than it solved". The difficulties pertained to its background assumptions and lack of empirical support. Seafloor spreading was highly suspect and geolarity reversals were still questionable. The two main groups working on reversal stratigraphy (A. Cox, R. Doell and G.B. Dalrymple of the US Geological Survey, and I. McDougall and D.H. Tarling at the Australian National University) were unwilling to commit themselves in print to the reality of field reversals until 1964, and the correct geolarity reversal time-scale was not established until the following year. The key empirical problems — failure to find (i) a ridge within the zebra pattern of magnetic anomalies in the north-eastern Pacific, and (ii) zebra patterns of the predicted magnetic anomalies surrounding known ridges — were respectively explained in 1965 and 1966.

Transform faults

Vine and J. Tuzo Wilson, then a recent and influential convert to drift, found the missing ridge in the north-eastern Pacific when Wilson and Hess visited Cambridge during the academic year 1964–1965. Central to their resolution was Wilson's concept of transform faults. (Like the Vine–Matthews–Morley hypothesis, the idea of transform faults was independently invented. Alan Coode came up with the idea in 1965 while a graduate student in the School of Physics at the University of Newcastle. His paper, like Morley's, was rejected by *Nature* and was published in an obscure journal where it went unnoticed.)

Wilson realized that seafloor spreading could be transformed into fault motion and thence into another spreading centre or trench system. The simplest examples of transform faults are the transverse fractures that offset ridge segments from each other. With the idea of transform faults in mind, Wilson and Vine noticed how an extension of the East Pacific Rise, the Juan de Fuca Ridge, ran through the zebra pattern of magnetic anomalies and connected with another section of the East Pacific Rise in the Gulf of California by means of the San Andreas Fault, which they identified as a transform fault.

They also recognized an additional aspect of the Vine–Matthews–Morley hypothesis, namely, that the pattern of

magnetic anomalies on both sides of an active ridge will be symmetrical if spreading rates are equal on both sides. The symmetry was obvious on the magnetic survey once they had located the axis of the Juan de Fuca Ridge. But their analysis had one problem — because they still were working with an incorrect timescale, they had to attribute an uneven spreading rate. This difficulty was resolved when Vine learned from Dalrymple in November 1965 that he, Cox and Doell had discovered two reversely magnetized examples dated at 0.71 and 0.72 million years, one of 0.88 million years having intermediate polarity, and a normally magnetized one of 0.89 million years. This enabled them to identify a period, which they called the Jaramillo, when the geomagnetic field had switched back to normal polarity. Before finding the reversed samples of 0.71 and 0.72 million years, they had thought that the first reversal had occurred around 1 million years ago. Upon learning of the Jaramillo, Vine became convinced of the correctness of his hypothesis and seafloor spreading, because the new timescale gave a constant spreading rate.

The second of the key empirical problems was resolved by researchers at Lamont-Doherty Geological Observatory. W. Pitman found that the Pacific-Antarctic Ridge was surrounded by a wonderfully symmetrical set of magnetic anomalies. Although unfamiliar with Vine and Matthews's original article, he recalled seeing a similar magnetic anomaly pattern in Vine and Wilson's discussion of the Juan de Fuca Ridge. Lamonters also came to realize that their survey of the magnetic anomalies surrounding the Reykjanes Ridge south of Iceland supported Vine's idea, contrary to

their original assessment. Opdyke, convinced of continental drift from his earlier work in palaeomagnetism, and one of the few at Lamont who supported it, found additional evidence for geopolarity reversals with his independent confirmation of the geopolarity timescale through his study of seafloor sediment cores — Opdyke, hired by Ewing to analyse the cores, began working on them after disagreeing with Lamont's initial interpretation of the magnetic anomalies surrounding the Reykjanes Ridge.

Meanwhile, L. Sykes, a seismologist, excited by Pitman's corroboration of the Vine-Matthews-Morley hypothesis, immediately confirmed Wilson's idea of transform faults from first-motion studies of earthquakes between ridge offsets. With the acceptance of seafloor spreading, Lamonters used their massive database to establish spreading rates for ocean floors.

Plate tectonics

Plate tectonics was independently conceived by Jason Morgan, a physicist turned geophysicist at Princeton, and Dan McKenzie, a recent PhD from the Department of Geodesy and Geophysics at Cambridge. Morgan, who shared an office with Vine (Hess invited Vine to Princeton, and Vine helped Morgan to understand Hess's ideas and the geological aspects of geophysics), presented his initial version of plate tectonics at the April 1967 meeting of the American Geophysical Union, but did not submit it for publication until August. After revision, it appeared in March 1968. Morgan, originally scheduled to speak about trenches at the AGU meeting, decided to present his more recent work on plate tectonics. McKenzie, having heard enough about trenches, decided to skip Morgan's talk, and therefore had the chance to invent plate tectonics himself. He came up with the idea while visiting the Scripps Institution of Oceanography. Afterwards, he asked R.L. Parker, who had taken a position there, for some technical advice about data presentation. Parker obliged with a computer program he had designed for plotting data on any map projection and their work was published at the end of 1967.

Plate tectonics grew out of the application of seafloor spreading and its two corollaries (the Vine-Matthews-Morley hypothesis and transform faults) to a spherical surface. Central to its development was the realization that Euler's theorem (which says that any movement of a rigid block on the surface of a sphere can be described by rotation about a point, or 'Euler' pole) could be used to describe the relative movements of around ten rigid plates comprising the Earth's outer layer, the lithosphere. Euler's theorem had already been used in connection with continental drift in 1965

JUAN DE FUCA RIDGE 46°N:130°W

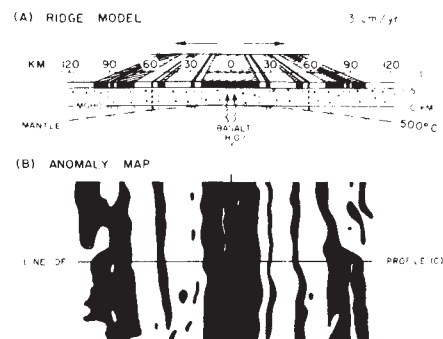


Fig. 3 The Vine-Matthews-Morley hypothesis combines seafloor spreading with recurrent reversals of the Earth's magnetic field, to explain the formation of magnetic anomalies. As the basalt of the newly created crust cools through its Curie temperature, it becomes permanently magnetized in the prevailing direction of the Earth's magnetic field. Crustal blocks magnetized in the same direction as today's magnetic field (shown in black above) are said to have normal polarity and give rise to positive anomalies in the magnetic profile. Reversely magnetized blocks (white) give rise to negative anomalies. In this way, knowledge of the history of magnetic reversals can be used to determine the spreading rate of a ridge. The line of the profile C is also shown (as C-C) in Fig. 1. (Vine, in Phinney, 1968.)

by Bullard, J.E. Everett and A.G. Smith in their computer-generated fit of continents surrounding the Atlantic ocean before they drifted apart — McKenzie came up with the idea while re-reading their article and thinking about transform faults.

Plate tectonics takes the rigidity of the moving sea floor, suggested by seafloor spreading and the Vine-Matthews-Morley hypothesis, extends it to the continents, and elevates it to an essential property of plates, for the rigidity of the plates allows the application of Euler's theorem. Morgan (who was also influenced by the work of W. Elsasser) and McKenzie found they could describe the motion of several plates in terms of the theorem, which meant that these plates moved as rigid bodies. X. Le Pichon, upon hearing Morgan's talk, returned to Lamont and extended the analysis to other areas of the world. McKenzie found evidence of rigidity through his analysis of heat flow and gravity data. Finally, B. Isacks, J. Oliver and Sykes presented extensive seismological evidence for the existence of the lithosphere.

Seismic activity is mostly restricted to the boundaries between plates, of which there are three types (see Fig. 2). Functionally defined in terms of plate creation, plate destruction or lack of either process, they are located at ridges (where plates are created), trenches (where plates are destroyed by subduction) or young fold mountains (where plates are deformed by thickening or folding), and transform

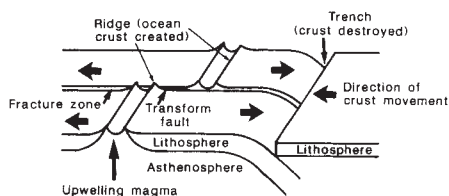


Fig. 2 In seafloor spreading, new ocean crust is created at mid-ocean ridges, where molten rock rises up from the mantle to fill the gap left by two diverging plates. The plate (or lithosphere) is made up of the ocean crust and uppermost mantle, which together act as a coherent and relatively rigid shell on top of the more fluid asthenosphere. The other end of the conveyor belt is the trench, or subduction zone, where the relatively cool, dense oceanic plate sinks back into the mantle. The process of subduction accounts for the fact that no sea floor older than Jurassic has been found. Mid-ocean ridges are divided into segments by transform faults, which offset the ridge axis, and hence the magnetic stripe pattern. Transform faults (and their 'fossil' traces, the fracture zones) can be used in plate tectonics to determine the direction of relative motion between two plates (see Fig. 4).

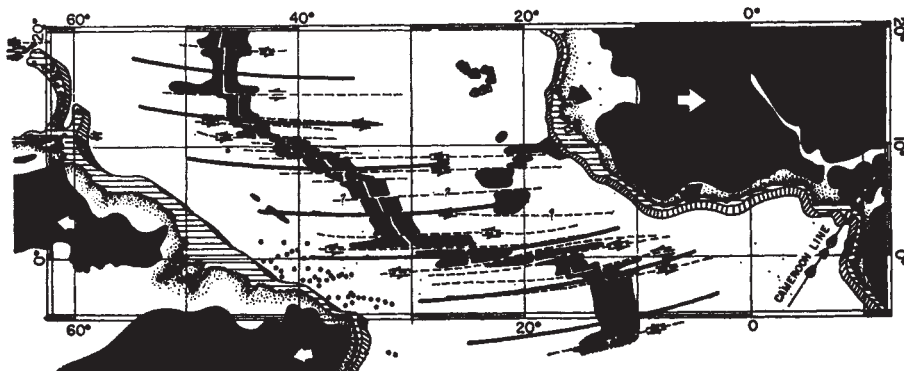


Fig. 4 Orientation of transform faults on the Mid-Atlantic Ridge can be used to locate the present pole of rotation of the African plate with respect to South America. Here the trends of fracture zones (dashed lines) are compared with small circles (solid lines) about an Euler pole at 62°N, 36°W. (After Morgan, 1968.)

faults (where plates slip past each other without being created or destroyed). Given Euler's theorem, the motion of one plate relative to another may be described as the rotation of one to the other about a pole on the surface of the Earth, where a full description of the motion requires location of the pole and the angular velocity of rotation about this pole. In addition, if adjacent plates are separated by transform faults, the faults are located on concentric circles about the pole of relative rotation. Thus, they define the direction of relative motion, and thereby acquire an added importance within plate tectonics which they did not have in sea-floor spreading (Fig. 4).

New framework

Plate tectonics provides Earth scientists with a new framework. Mobile plates, made up of rigid lithosphere, replace static continents and oceans as the key components in the evolution of the Earth's surface features. And the important boundary between the outermost layers of the Earth is not between crust and mantle, but further down, where the rigid lithosphere is separated from the weaker asthenosphere (see Fig. 2).

The revolution explains why marine Earth scientists have learnt more, much more quickly, than have their counterparts investigating the continents, for we now know that ocean basins live much shorter and less eventful lives than continents. Because continents are more buoyant than the sea floor, they remain on the Earth's surface where they move about, collide, break apart and form new aggregates. Sea floor is created by a process which is not obscured by interfering events, and it does not deform as it migrates away from ridge axes until subducted or welded onto a continent. The immediate success of plate tectonics, enhanced by McKenzie and Morgan's analysis of the evolution of triple junctions (places where three plates meet), was its detailing the evolution of ocean basins — their histories, explicable in terms of plate movements, even included postulation of

plates that no longer exist.

This quantitative description of the history of ocean basins and controlling processes, plate tectonics' crowning success, has had a considerable effect upon historical geology. The distrust or indifference of many geologists during much of this century towards causal theories of global scope and geophysical studies was partly due to their view that such theories and findings had little success explaining, in anything but general terms, the Earth's geological history. But the geochronology recorded by marine magnetic anomalies, echoed in deep-sea sediment cores, and explained by seafloor spreading and geolarity reversals, has given Earth scientists a better continuous chronicle than ever recorded on the continents.

The continental-drift aspect of plate tectonics was also quickly exploited by geologists, palaeoclimatologists and biogeographers. Besides applying continental drift to problems in their speciality, they offered — often in conjunction with palaeomagnetists — pre-Jurassic reconstructions and traced geological matches between formerly joined continents.

Much of the heuristic value of plate tectonics is found in the unified framework it offers Earth scientists in dealing with regional problems, for plate tectonics has led to the identification of new problems and new ways of analysing old problems. Realizing that, in plate tectonics, areas of plate creation and destruction are restricted to creative (divergent) and destructive (convergent) boundaries, researchers have concentrated upon them, asking questions about the micro-processes involved in plate formation and destruction. Convergent boundaries have received considerable attention, leading to new theories of mountain building and island-arc formation as well as the theory of exotic terranes. Also, once palaeomagnetists showed that the continents had drifted before Wegener's Pangaea was formed — the vindication of palaeomagnetists is another result of the plate tectonics revolution — Earth scientists have looked for evidence of former collisions

between continental sections of plates. This has opened up a nest of new problems in pre-Jurassic geology, and is one way plate tectonics has been applied to continents. By treating suture zones as fossilized destructive plate boundaries, geologists and palaeomagnetists have applied plate tectonics to areas within continents which were once continental margins. However, what remains problematic for plate tectonics, as Molnar explains in the following Review Article, is its application to continental interiors where deformation is currently occurring.

Plate tectonics has shown that the question "How can the continents plough through a rigid sea floor?", the nemesis of continental drift, is not the question to ask. Rather, the key dynamical question is "What forces drive the plates?". Although some progress has been made, the issue remains unsolved. But, unlike continental drift, plate tectonics has not been rejected because the dynamical problem remains unsolved, partly because plate tectonics, unlike drift, is viewed only as a kinematic theory. Also, because the empirical support for plate tectonics provided by confirmation of Vine-Matthews-Morley and transform faults is so much stronger than what was available for continental drift, the impossibility or improbability of plate tectonics has never been an issue. Moving plates, like evolving species, have become accepted as matters of fact. Their cause, however, remains open to dispute.

Even though most geologists probably still believe that most geophysicists cannot tell a rock from a hard place, and most geophysicists think that most geologists cannot balance their chequebooks, relations have improved between the two groups in part because of the revolution in the Earth sciences. Seismologists and palaeomagnetists, in particular, have had tremendous success in dealing with structural issues — the latter working on present-day movements, the former on past ones.

In 1951, the young geophysicist, Runcorn, had the good sense to hire Irving, a young geologist, and Irving had the good sense to learn about geomagnetism. Much has happened since they found that their palaeomagnetic work was more welcome among members of the Royal Astronomical Society than the Geological Society of London — the cooperation among palaeomagnetists with backgrounds in either geophysics or geology has served as a model for how properly to investigate problems in the Earth sciences. □

Background research has been supported by NSF and NEH. I thank K. Creer, J. Clegg, J. Hospers, E. Irving, D. Matthews, D. McKenzie, L. Morley, N. Opdyke, S.K. Runcorn and F. Vine for answering questions about their work; T. van Andel, P. Molnar, A.G. Smith and F. Vine for helpful comments on earlier drafts; and N. Biersmith for numerous retypings of the article.

Henry Frankel is in the Department of Philosophy, University of Missouri — Kansas City, Kansas City, Missouri 64110, USA.

Continental tectonics in the aftermath of plate tectonics

Peter Molnar

Department of Earth, Atmospheric and Planetary Sciences, MIT, Cambridge, Massachusetts 02139, USA

The success of plate tectonics required an acceptance of continental drift, and thus a reinterpretation of the large-scale geological history of most of the Earth. But the basic tenet of plate tectonics, rigid-body movements of large plates of lithosphere, fails to apply to continental interiors, where buoyant continental crust can detach from the underlying mantle to form mountain ranges and broad zones of diffuse tectonic activity.

TWENTY-FIVE years ago, to explain some subtle, almost insignificant spatial variations in the Earth's magnetic field, Vine and Matthews¹ integrated two wholly unproven and otherwise unrelated hypotheses—the creation of new sea floor at mid-ocean ridges and recurrent reversals of the Earth's magnetic field. In doing so they helped launch a revolution in the Earth sciences. Reduced to its essentials, plate tectonics is the quantitative description of the kinematics of the lithosphere, the strong outer layer of the Earth including the uppermost mantle and its overlying crust. Vine and Matthews presented a method for measuring how rapidly plates diverge from one another at oceanic ridges, or spreading centres. Because most plate boundaries are very narrow, often consisting of only one fault or fault zone, accurate mapping of the bathymetric expression of these zones and the study of earthquakes at these boundaries allowed the directions of relative motions between neighbouring plates to be determined. By the late 1960s, geophysicists had discovered an internal consistency of the rates and directions of relative motions of these vast aseismic plates, at least where the boundaries between them are indeed narrow, and therefore had demonstrated that the plates move as effectively rigid bodies over the surface of the Earth. This is plate tectonics.

Plate boundaries in oceanic regions are narrow. In contrast, seismicity and active faulting in continental regions are commonly dispersed over thousands of kilometres. Today, one of the central controversies among students of continental tectonics concerns how best to describe this widespread deformation: by the relative motion of rigid blocks ('platelets'), or by the deformation of a continuous medium that undergoes internal strain and rotation (or vorticity). This is not to say that plate tectonics has failed to help us understand how continents deform or evolve, but that the fundamental assumption of plate tectonics, rigid plate motion, provides an incomplete description of the broad zones of deformation within continents.

Growing appreciation for this limitation has defined the study of continental deformation for the last fifteen years, but not until after attempts had been made to apply plate tectonics to continental regions. As marine geologists established plate tectonics as the paradigm for the tectonics of oceanic terrains, land-based geologists were quick to realize that the geological record contained evidence for plate motions earlier in the Earth's history. For example, they recognized that the mélanges of oceanic sedimentary rock, basalt, and ultrabasic rock, typical of oceanic crust and mantle (such as those found along the coast of northern and central California), and the parallel belts of granite (cropping out in the Sierra Nevada in eastern California), constitute geological evidence for the underthrusting and off-scraping of oceanic floor beneath the granitic belt^{2,3} (east-

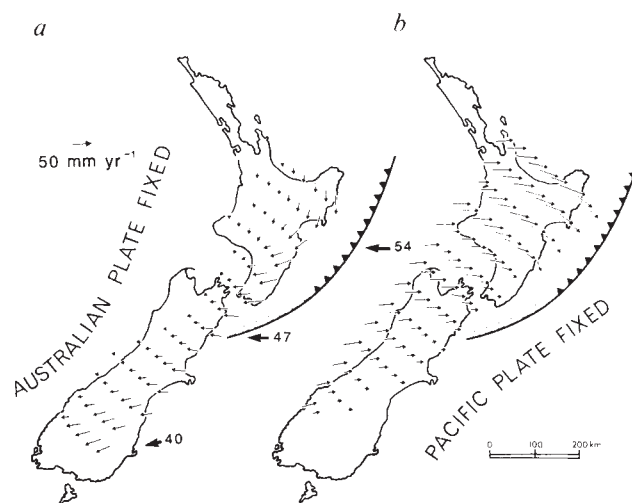


Fig. 1 Maps of New Zealand (after ref. 6) showing velocity fields over the islands relative to the Australian (a) or Pacific (b) plate. Note diffuse simple shear across both islands, as well as subduction of Pacific Ocean floor beneath the North Island perpendicular to the Hikurangi Trough (barbed line). The three large arrows indicate the relative velocity between the Australian and Pacific plates, in mm yr^{-1} .

ward beneath western North America). A present day analogue is the subduction of ocean floor beneath the volcanically active Andean margin of South America. Such 'geological corollaries' for processes clearly occurring elsewhere today not only allowed but required a reinterpretation of the geological histories of vast areas, such as the western United States. Regional synthesis became a respectable domain of serious scientists, a status that had been lost since the very perceptive insights of geology's grand thinkers, Wegener, Seuss and Argand, 50 to 100 years ago.

Plate motions and continental deformation

The acceptance of continental drift, seafloor spreading and subduction of oceanic crust—ancestral concepts linked together by plate tectonics—required a qualitative reinterpretation of continental geology, but the application of the quantitative aspects of plate tectonics revealed both the full power of plate tectonics applied to continental regions, and its limitations. The Vine-Matthews hypothesis applied to the sea floor of the Pacific west of North America enabled Atwater⁴ to determine its age

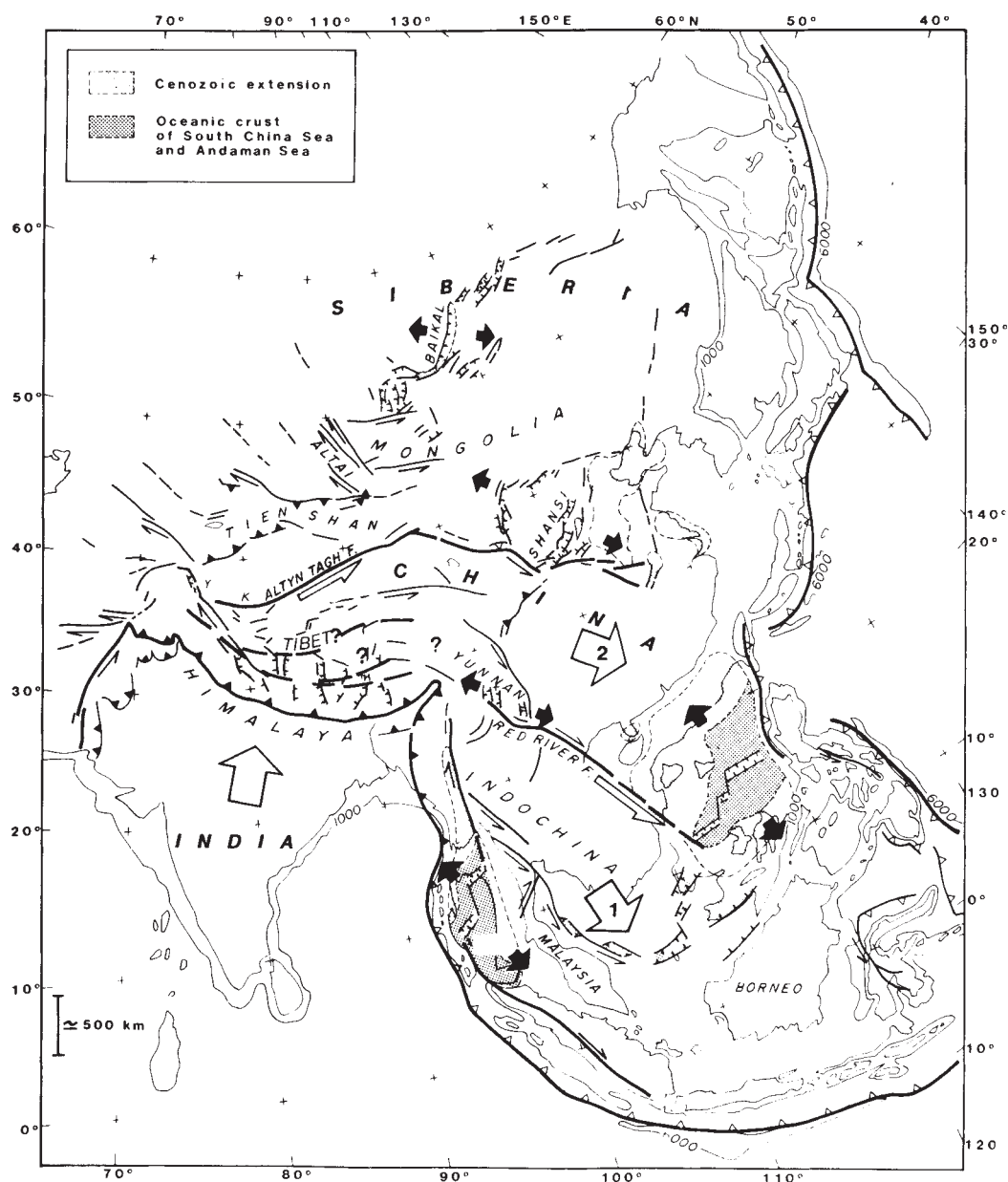


Fig. 2 Map of active faults and tectonics of Asia (after refs 8, 18). Northward penetration of India towards Siberia causes thrust faulting (lines with barbs) at the Himalaya and Tien Shan, and strike-slip faulting, primarily left-lateral, on faults emanating from Tibet and across Mongolia, as far north as Baikal. Extrusion of material sideways from in front of India's path may have displaced Sundaland and south-east China to the south-east^{7,8,17,18}. In a thrust fault, one block of crust is pushed over another; on a strike-slip fault, blocks of crust slide past each other, moving horizontally.

and evolution and then the implications of that evolution for the geological history of western North America. Not only had oceanic crust been thrust beneath California, but that process had continued until much more recently than could be inferred from ages of rocks either along the coast or in the Sierra Nevada. Atwater also showed that slip on the San Andreas fault in southern California, initially considered to be the plate boundary between North America and the Pacific plate, has been much smaller than the displacement of those two plates since they came in contact. Most significantly, her work emphasized that an understanding of tectonic processes required much more than the analysis of rock exposed at the Earth's surface. Although Atwater's application of the tenets of plate tectonics to the oceanic realm in order to study the timing and scale of processes within continental North America forced a reassessment of the geological evidence for the tectonic evolution of California, it provided little insight into how the interior of western North America had deformed. She simply wrote this off as deformation of a "wide, soft zone"⁴.

The interrelationship between rigid motion of oceanic litho-

sphere and more diffuse deformation in continental regions is particularly clear in New Zealand, the islands of which straddle a part of the boundary between the converging Pacific and Australian plates. The Pacific plate, which includes the southeastern part of New Zealand, rotates anticlockwise with respect to Australia, the Tasman Sea and the rest of New Zealand about a point ~1,000 km south of New Zealand. Pacific ocean floor is underthrust beneath the North Island of New Zealand, but deformation occurs throughout the crust of both islands (Fig. 1). Re-surveys of bench marks installed over much of the country in the last century revealed relative displacements of the bench marks that could be described only as diffuse, inhomogeneous strain. Nevertheless, an integration of that strain across the South Island yields a rate and direction of convergence of its margins that agrees remarkably well with the velocity calculated using the Vine-Matthews hypothesis and data from spreading centres in the South Pacific and the Indian Ocean^{5,6}. The agreement is a tribute both to plate tectonics and to surveying techniques developed in the last century, but perhaps the most important result is the need to describe the deformation

in New Zealand in terms of strain and vorticity instead of rigid-body motion. Moreover, the internal rotation (or vorticity) deduced from the re-surveys is corroborated by rotations, measured using palaeomagnetism, of small areas within New Zealand⁶. Thus, although one might conceivably be able to describe the deformation of New Zealand as relative displacements and rotations of numerous small blocks of crust, the number and the dimensions of such blocks would probably make such a description impractical.

The region where intracontinental deformation occurs on its grandest scale is Central Asia (Fig. 2). The penetration of the Indian subcontinent into the rest of Asia seems to be responsible for the world's highest mountains (the Himalaya, a consequence of the overthrusting of slivers of India's northern margin onto the rest of the subcontinent) and the largest and highest plateau (Tibet, a consequence of apparently distributed, crustal thickening), as well as the world's deepest lake (Baikal, filling a rift system caused by a splitting apart of the Asian landmass)^{7,8}. Within the area of active deformation, ~3,000 km across, are large regions, such as the Tarim basin, that do not seem to be undergoing any deformation, and regions of comparable dimensions, such as the Tibetan Plateau, that are sliced by numerous faults, which isolate many tens of separate small blocks. The deformation that we see at the Earth's surface is neither homogeneous nor accurately described by the relative motion of a few rigid blocks.

Thus, whereas the success of plate tectonics on a global scale encouraged Earth scientists to analyse areas as large as Central Asia in terms of general processes, and whereas many of the same techniques that contributed data fundamental for establishing plate tectonics were also vital for obtaining a qualitative image of how deformation occurs in continents, we still lack a quantitative description of the kinematics of this deformation comparable with that used to describe plate motions in oceanic regions. The broad, diffuse deformation of the western United States, of New Zealand, or of Central Asia is much more complex than the rigid-body displacements of a small number of large plates, and finding a simple and accurate way to represent the deformation of continents remains a major task.

In short, plate tectonics is a poor approximation for the tectonics of many continental regions because vast portions of continental crust and mantle do not seem to move together as parts of the same, coherent rigid plates. Why should this be?

Differences between continents and oceans

Oceanic lithosphere has only a few kilometres of crust overlying its more dense mantle portion, but continental lithosphere commonly includes 35 km of relatively light crust. The difference is crucial: whereas the thin oceanic crust appears to be dense enough to plunge back into the mantle at subduction zones, the buoyancy of thick continental crust keeps it afloat. If continental lithosphere were strong enough to maintain its integrity at a subduction zone, the buoyant continental crust would not only resist being subducted, but the subducting plate would abruptly grind to a halt when the continental 'passenger' reached the trench.

The strength of continental lithosphere also contrasts with that of oceanic lithosphere. The strongest part of oceanic lithosphere seems to lie in the mantle, between 20 and 60 km depth, beneath a brittle upper part and above its increasingly ductile lower part, which grades downward into the asthenosphere (Fig. 3). In the same depth range where oceanic lithosphere is strongest, however, continental lithosphere consists of crust, not mantle. At temperatures typical of the lower crust (400–700 °C), the minerals comprising the crust appear to be much weaker than olivine, the strong mineral that comprises most of the upper mantle. Consequently, continental lithosphere could be much weaker than oceanic lithosphere. Oceanic lithosphere behaves as a virtually rigid plate because of its strong core, but, as the late C. Goetze noted in the mid-1970s, continental lithosphere

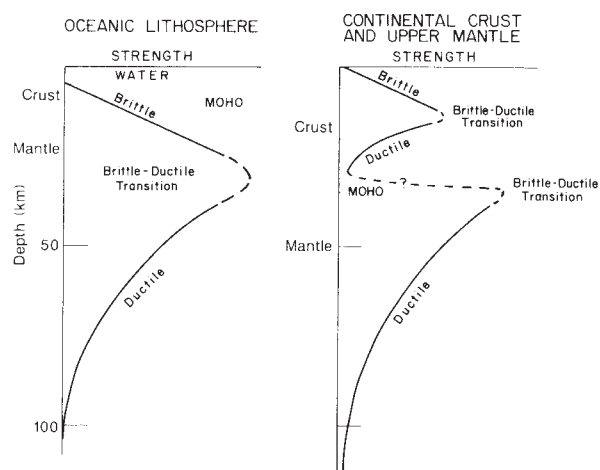


Fig. 3 Profiles of strength across oceanic and continental lithosphere, based on frictional strengths at shallow depths and ductile flow laws of minerals at greater depths⁹. Note the strong core of oceanic lithosphere at depths of 20–60 km and the probable weak, ductile region in the lower continental crust. Obviously, other phenomena, such as fluids, impurities in minerals or melting, can modify these curves, and given the uncertainties in the flow laws, the strengths cannot be usefully quantified. The existence of a low-strength zone in continents, however, seems very likely. The boundary between crust and mantle, known as the Moho, is defined chemically; mantle rocks are much richer in magnesium and iron than either continental or oceanic crust. By contrast, the boundary between lithosphere and asthenosphere is defined by a change in rheology; the lithosphere comprises the crust and uppermost upper mantle, which together make up a relatively strong shell overlying the weaker asthenosphere.

might consist of three layers: a brittle upper-crustal layer, a weak lower crust and a stronger uppermost mantle, which, nevertheless, would not be as strong as the strongest part of the oceanic lithosphere⁹. This jam-sandwich-like rheological profile (Fig. 3) is also suggested by the frequent occurrence of earthquakes (brittle deformation) in the upper crust, their nearly complete absence in the (presumably weak, ductile) lower crust, and their occasional presence in the underlying upper mantle¹⁰.

These differences in rheology and buoyancy are probably responsible for mountain ranges, the most spectacular features of the continents, and for their virtual absence in oceans. (Mid-ocean ridges and seamounts owe their relief primarily to volcanic constructional processes and elevated temperatures in the underlying mantle, rather than to 'mountain building' as described below.) Most mountain ranges are built by crustal thickening. When masses of crust are pushed together, their buoyancy inhibits subduction into the mantle, and slices of upper crust, detached from the region below, are thrust atop one another. Crustal thickening does not seem to occur in oceanic regions (where, instead, subduction of the entire lithosphere allows sustained convergence of two lithospheric plates), but it dramatically affects the style and amount of deformation in continents.

For crustal thickening to occur, and for mountain ranges to be built, work must be done against gravity. Mountain ranges and high plateaux store gravitational potential energy, and this potential energy varies with the square of the mean height¹¹. A rapidly increasing amount of work must be done to increase the mean height of an already elevated region, and therefore high plateaux created by crustal shortening and thickening should reach limiting elevations^{11,12}. Continued convergence will then manifest itself by a growth in area of the range or plateau, as thrust faulting leads to crustal thickening on its margins. Hence, a high plateau can serve as a 'pressure gauge' for the average pressure applied to its margins, and it can transmit this pressure, much as a contained fluid transmits pressure laterally across

itself^{11,12}. Note the important difference between the sustained, virtually steady-state subduction of oceanic lithosphere into the asthenosphere and the transient growth of continental plateaux, by a process that has no analogue in plate tectonics.

Kinematics and dynamics

A key to the simplicity of plate tectonics is that the strength of lithospheric plates enables the analysis of their kinematics to be isolated and treated separately from the dynamic processes controlling plate motions; relative velocities of plates can be analysed without reference to the forces that give rise to them. Conversely, the much different rheological structure of the continents, which allows rapid deformation of the crust and mantle and facilitates the decoupling of one from the other, leads to an inseparable dependence of the kinematics of continental deformation on the dynamic processes controlling that deformation. In addition, because of crustal thickening, deformation varies in all three dimensions, and horizontal variations in mean elevation and in crustal thickness require horizontal gradients in some components of stress. Thus, not only do we lack a well-defined procedure for describing the kinematics of continental deformation, but we must also recognize the ultimate necessity of considering the dynamics in three dimensions.

We cannot do this yet, and even progress in understanding the dynamics of continental deformation in two dimensions is recent. In an important study, England and McKenzie¹³ analysed the deformation of a thin viscous sheet as an analogue for the continental crust and upper mantle. By considering vertically averaged stresses and strains, they could study two-dimensional variations in the deformation of such a sheet in a gravity field. Calculations of stress, strain and vorticity fields and of distributions of equivalent elevations and crustal thicknesses as a function of time depend on only two non-dimensional numbers; one, the Argand number, measures the relative importances of gravity and of material strength in inhibiting deformation, and the other is a measure of the nonlinear dependence of strain rate on stress. With only two non-dimensional parameters, a full series of numerical experiments could examine the dependences of the calculated fields on them and on the boundary conditions. The most significant result of these numerical experiments is the weak dependence of the deformation field on the geometry of the boundary conditions; deformation of a sheet indented by a rigid die decays rapidly at distances from the die comparable to its width^{14,15}. By analogy with the Earth, England and Houseman¹⁶ inferred that the penetration of India (a rigid die) into Eurasia (a thin viscous sheet) during the last 40–50 Myr caused crustal thickening in the area ~1,000 km north and north-east of the Himalaya, but that south-east China has not been extruded very far to the east, out of India's northward path (see Fig. 2).

In another approach, laboratory studies by Peltzer and Tapponnier^{17,18} of the indentation of plasticine, unconfined on its top and bottom and on two sides but unconfined on one lateral edge, addressed the rapid eastward extrusion of parts of China and Indochina, which had been inferred from the prevalence of strike-slip faulting in Asia^{7,8}. They found a large lateral extrusion of coherent, if deformed, pieces of plasticine^{8,17,18}. The small expulsion calculated for the thin viscous sheet and the large expulsion of plasticine are largely a consequence of thickening of the indented viscous sheet but not of the plasticine, but the difference between the rheological properties (or constitutive laws) of the two media also contributes to the differing observed styles of deformation.

Role of crustal blocks

This difference in assumed constitutive laws underscores one of the principal questions asked about the dynamics of continental deformation. The thin viscous sheet deforms homogeneously, but plasticine undergoes strain softening, with the localization of shear in narrow zones, analogous to faults and fault zones

in the Earth. The concentration of deformation in shear zones allows intervening regions to be displaced coherently, as blocks (or pieces of plasticine), but homogeneous deformation does not include blocks that translate with respect to one another. Clearly, the crust that we see and walk on is not flowing or deforming with large strains, and the issue, therefore, is not whether or not there are blocks of crust at the surface, but what role they play.

In one extreme, the upper crust is imagined to consist of decoupled blocks of rock whose composite relative displacements serve primarily as passive markers that reflect the average deformation of ductile material below^{6,13–16,19,20}. Hence, the velocity or displacement fields of points on the surface of the Earth would yield a rough image of ductile deformation of (possibly stronger) material at depth. In the other extreme, the outer part of the Earth deforms by localization of strain into strong, if deformable, blocks, whose relative motions are controlled by stresses on their edges^{8,17,18,21}. Effectively, such blocks float on a weak substratum, the asthenosphere, in a manner similar to ice fragments floating on water. In the former case, slow motion of the underlying fluid carries the blocks, and in the latter, boundary conditions on the lateral margins of the deforming continent form blocks that rub on one another and move above a passive fluid. Because studies of the viscous sheet ignore localized weaknesses, and because the strongest part of continental crust and upper mantle is thought to lie just below the Moho⁹, the thin viscous sheet approximates the former more than the latter analogue. A medium, like plasticine, cut by prominent shear zones is more analogous with the latter.

Clearly, these two oversimplified images of crustal blocks driven by flow beneath them or by stresses on their margins represent end-members which embrace the likely relationship between the observable kinematics of continental deformation and the dynamic processes that we try to infer. Consideration of them as extremes, however, allows us to pose very simply a number of specific problems that should reduce our ignorance of how the dynamics and kinematics are related.

For example, suppose the thin viscous sheet were an appropriate analogue for continental tectonics, and specifically suppose that England and Houseman¹⁶ were correct in suggesting that most of the extrusion of Tibet due to India's penetration into Asia has been absorbed by crustal thickening on the eastern edge of Tibet and not by southeastward expulsion of south-east China. Then, slip on the major strike-slip faults on the margins of Tibet and south-east of it should be small, less than 100 km. Alternatively, suppose that Peltzer and Tapponnier^{8,17} were correct that the deformation is controlled by the localization of slip on major strike-slip faults and shear zones, and, in particular, that expulsion of blocks comprising Indo-China and South China has been large. Then, the cumulative slip on several faults should also be large, more than 100–300 km on each fault, summing to at least 1,000 km of cumulative displacement. Burchfiel's²² and Zhang Peizhen's²³ measurement of only 10–15 km of Cenozoic slip on the Haiyuan fault, the only significant strike-slip fault strand in China with a well constrained total offset, does not imply a large expulsion, but clearly the displacement on only one fault is insufficient to prove this. In any case, classical geological techniques for mapping offsets on major faults hold the key to resolving this question, and quantifying the amounts of deformation or displacement looms as one of geology's great challenges for the future.

Although geological methods are honed for mapping both displacements of coherent rock units at faults and large internal strain of individual terrains, rotations of small regions or blocks about vertical axes can easily go undetected; yet such rotations may hold a key for understanding the relation between surficial and deeper deformation. Beginning largely from the work of Beck²⁴ and Luyendyk²⁵, a wealth of palaeomagnetic data now show large, late Cenozoic rotations of isolated crustal blocks^{24–28}, and such rotations comprise an important part of

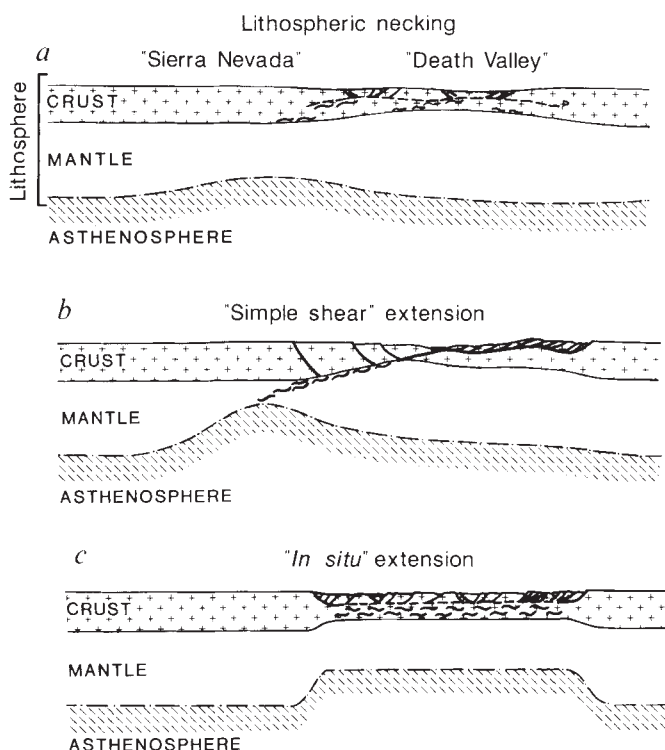


Fig. 4 Idealized cross-sections across the Basin and Range Province and the Sierra Nevada in California (after ref. 38), illustrating the likely possibility (*a, b*) that crustal extension in Death Valley may be detached from the mantle directly underlying it, and the extension of the mantle lithosphere may underlie the Sierra Nevada. The youthful uplift of the Sierra Nevada would be a consequence of the intrusion of hotter material where the underlying mantle lithosphere had thinned. Cross-section *c* shows the older, apparently incorrect idea that the crust and its directly underlying mantle would be stretched and thinned by the same amount.

the deformation in broad shear zones^{6,29,30}. If such blocks were dragged along (like ball bearings) by horizontal displacements at the edges of the shear zone, the rate of rotation would be simply the rate of slip of the margins of the shear zone divided by the radius of the block. If, instead, the block floated on a viscous medium undergoing distributed homogeneous shear, the rate of rotation would be half as large (because only the vorticity, and not the pure shear, of the viscous material would contribute to the rotation^{19,20}). Thus, rates and amounts of rotations of blocks driven by stresses on their edges or by flow beneath them differ from one another. To distinguish between these two extremes requires accurate measurements both of displacement of opposite sides of the shear zone and the rotation of the intervening material, as well as a belief that the areas that rotate are equidimensional³¹. What measurements exist (see, for example, refs 6, 20, 28, 32) favour blocks floating on a viscous substratum undergoing shear, but whether this process is common or not is not known.

Detachment of crustal fragments

If blocks were driven by stresses on their margins, then boundaries between them would probably pass directly from the Earth's surface through the stronger mantle beneath them. If they were driven by flow beneath them, then the boundaries between such blocks might be truncated by flat shear zones in the crust. One of the exciting discoveries of the 1980s is the widespread detachment of small blocks of the upper crust from the lower crust and uppermost mantle. Following the work of Anderson³³ and Armstrong³⁴, field geologists working in the western United States have mapped abundant evidence for large

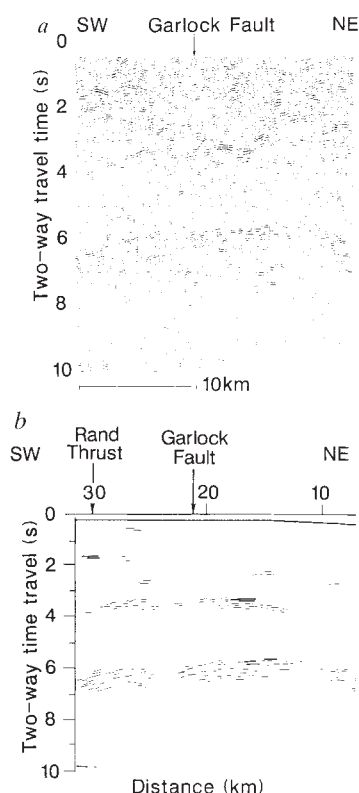


Fig. 5 Seismic reflection profile across the Garlock Fault in California (*a*) and simple interpretation of it (*b*). The Garlock Fault is a major strike-slip fault, with many tens of kilometres of displacement³⁹, but the fault does not seem to penetrate below ~10–12 km depth (two-way travel time of 3.5 s). Instead, nearly flat reflections pass beneath the surface trace of the fault and could mark detachment of the upper from the lower crust. (From refs 40 and 41.)

horizontal translations of highly fractured blocks of upper crust over ductilely sheared, strongly metamorphosed (obviously hotter) lower-crustal rock. Whether the contacts between these brittle deformed upper-crustal rocks and ductilely deformed lower-crustal rocks mark single faults³⁵ or a more complicated juxtaposition of two entirely different styles of deformation^{36,37} is unresolved, but it seems likely that the stretching of the crust in one area might continue into the mantle elsewhere³⁸ (Fig. 4).

A lack of coherence between the upper continental crust and its uppermost mantle is also virtually certain along some strike-slip faults. Where strike-slip faults mark the boundaries between oceanic plates (where they are known as transform faults), the rigidity of the plates requires that the rate and amount of displacement be the same at all points along the fault. In continents, however, the deformation of material on both sides of the fault can be so large that the amount of strike-slip displacement varies markedly along the fault³⁹. If this deformation involved the entire crust, there would be marked differences in crustal thickness across the fault. Seismic imaging of some intracontinental strike-slip faults, however, suggests that these faults do not penetrate deeper than 8–10 km^{40,41} (Fig. 5). Even portions of the San Andreas fault in California may not continue through the upper mantle directly beneath it⁴². Although the surface trace is so overwhelmingly clear that pilots flying between Los Angeles and San Francisco can navigate by it, part of the mantle lithosphere beneath the Transverse Ranges seems to detach and plunge into the asthenosphere instead of being displaced coherently with the crustal blocks on either side of the fault⁴³. Elsewhere, the San Andreas fault seems to pass directly into the mantle as a narrower vertical zone⁴¹. Thus,

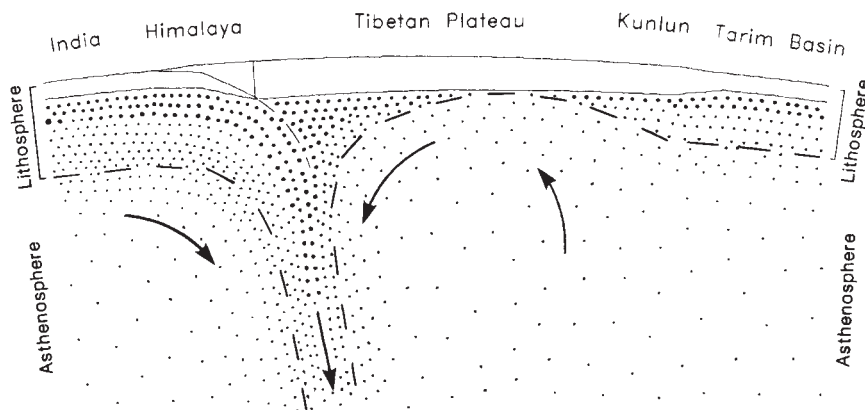


Fig. 6 Inferred dynamics of the mantle beneath Tibet and the Himalaya (from ref. 47). Low seismic-wave velocities in northern Tibet may mark a zone of upwelling in the asthenosphere. The flexure of the Indian lithosphere beneath the Himalaya seems to require a bending moment applied to it, and convective downwelling beneath southern Tibet could be the source of that moment⁴⁶. Large, closely spaced dots indicate colder temperatures, grading into hotter asthenosphere with smaller, widely spaced dots.

small crustal blocks separated by minor faults probably are detached from the lower crust and upper mantle beneath them, but major faults may pass through the strong uppermost mantle as localized shear zones which separate plates, or small fragments, of intact continental lithosphere.

Future directions

Ultimately, a full understanding of both the kinematics and the dynamics of continental tectonics will include aspects of both a continuous medium and the movements of strong blocks. The question is how small blocks must be before their relative motions are more accurately described as parts of a continuum. At one extreme, a crustal block the size of Siberia is a plate, as is the Tarim Basin, which at $500 \times 1,000$ km behaves as a large rigid block surrounded by zones of diffuse deformation. The existence of only a few strike-slip faults in Asia along which rates of slip exceed 10 mm yr^{-1} also implies a concentration of deformation in a few zones separating extensive, less rapidly deforming regions, 200–500 km across. Near the other extreme, each range in the Basin and Range Province of the western United States, ~ 10 –20 km across, seems to move with respect to the others. A description of the velocity field for the Basin and Range Province will surely require many more parameters than the four elements of a two-dimensional velocity gradient tensor. Moreover, if these small blocks are indeed detached from the underlying lower crust and mantle, the deformation of this deeper region could be continuous, and perhaps relatively homogeneous. Hence some aspects of continental deformation can be described well by relative motions of relatively rigid blocks, which behave as small lithospheric plates, but others seem to require the consideration of a continuous medium. Like all complex questions, this becomes one not only of kind, but also of degree.

Thus, one focus of future studies of continental tectonics will surely be on measuring the kinematics of deformation. Traditional geological field mapping applied to quantifying deformation will continue to flourish, and its burgeoning offspring, the study of Quaternary faulting, will surely yield the most precise average velocities of slip on faults. The task of determining rotations, however, will fall largely on the palaeomagnetists. Because deformation in the outer part of the Earth occurs largely by slip during earthquakes, seismology will always hold a seat in the presidium of continental tectonics. But the most rapid advances in our understanding of the kinematics of continental deformation may come with the imminent renaissance in surveying, brought about by the implementation of Global Positioning System satellites, which will make repeated measurements of relative positions of points tens to hundreds, and even thousands of kilometres apart easy, and eventually inexpensive. One can foresee a time when the deformation of continents will be monitored continuously.

I have focused on the importance of the kinematics of deformation as crucial for understanding why and how continents deform as they do, but obviously such studies will not provide all the answers. Two examples illustrate this. First, careful laboratory work led Goetze and his colleagues⁹ to recognize the likely difference in profiles of strength through continental and oceanic lithosphere (Fig. 3). Second, imaging of the deep structure of continental regions, largely with seismic waves (see, for example, refs. 40–43), is vital for understanding how and where the upper crust detaches from the underlying material—for example, on narrow flat faults or a broad shear zone.

Back to dynamics

The tectonics of continents has found plate tectonics an inadequate paradigm. Nonetheless, it shares one unsatisfactorily answered and profoundly important question: What is the driving mechanism?

Plate motions are the kinematic manifestations of large-scale convection in the Earth; oceanic lithosphere is the cold boundary layer of that convection. It is a simple matter to ascribe the driving force to gravity causing plates to slide downhill from mid-ocean ridges and pulling them into the asthenosphere at subduction zones, but it is a rare fluid dynamicist who would contend that these processes are understood or that such a description constitutes a physically rigorous, quantitative description of the driving mechanism. Similarly it seems probable that mountain belts, at which crustal shortening is a major process, overlie zones of convergence and downwelling convective flow in the mantle^{43–47} (Fig. 6). What is not known about the structure of the upper mantle beneath major mountain belts, however, dwarfs what is thought about such regions. A growth in our knowledge of lateral variations in the structure of the upper mantle is vital for an understanding of convection in the mantle and the forces driving plate motions, which will, in turn, bring a refined understanding of the forces that drive mountain building and continental tectonics.

The contribution of plate tectonics to continental tectonics has been enormous. The demonstration of large horizontal displacements of terrains over the surface of the Earth required not only the consideration of large-scale processes on the continents but also a very different understanding of the geological record than had previously been widely held. The most profound effect of plate tectonics on the Earth sciences, however, was to expedite its transition from a largely qualitative, data-cataloguing natural science to a quantitative physical science. One obvious indication of this has been the enormous growth of geophysical methods for solving geological problems—those same methods that had demonstrated plate tectonics. A second is the growing emphasis on problem solving. Most geologists now map regions, not to put colour on previously blank maps, but rather to solve specific problems that require better geologi-

cal maps than currently exist. Similarly, marine geologists no longer ply the seas in a spiderweb of ship tracks, but now choose specific areas for detailed study of particular phenomena. Perhaps the most rewarding and the most useful result of plate tectonics was the recognition that processes that had seemed too difficult to study except qualitatively could, in fact, be analysed quantitatively with simple physical concepts replete with mathematics and meaningful uncertainties. Although continental tectonics has not had a revolution like plate

tectonics, it has nevertheless progressed enormously in the past 20 years.

I am grateful to B. C. Burchfiel, P. C. England, J. A. Jackson, R. McCaffrey, D. P. McKenzie, S. Neustadt and P. Tapponnier for critically reading earlier drafts of this manuscript and D. L. Frank for numerous retypings of it. R. Allmendinger, C. H. Jones, S. Kaufman, P. Tapponnier and R. I. Walcott kindly supplied original figures. I thank NASA and NSF for continued support.

1. Vine, F. J. & Matthews, D. H. *Nature* **199**, 947-949 (1963).
2. Hamilton, W. *Geol. Soc. Am. Bull.* **80**, 2409-2430 (1969).
3. Dewey, J. F. & Bird, J. M. *J. geophys. Res.* **75**, 2625-2647 (1970).
4. Atwater, T. *Geol. Soc. Am. Bull.* **81**, 3513-3536 (1970).
5. Bibby, H. M. *Geophys. J. R. astr. Soc.* **66**, 513-533 (1981).
6. Walcott, R. I. *Geophys. J. R. astr. Soc.* **79**, 613-633 (1984).
7. Molnar, P. & Tapponnier, P. *Science* **189**, 419-426 (1975).
8. Tapponnier, P., Peltzer, G. & Armijo, R. in *Collision Tectonics* (eds Coward, M. P. & Ries, A. C.) 115-157 (Geological Society, London, 1986).
9. Brace, W. F. & Kohlstedt, D. L. *J. geophys. Res.* **85**, 6248-6252 (1980).
10. Chen, W.-P. & Molnar, P. *J. geophys. Res.* **88**, 4183-4214 (1983).
11. Molnar, P. & Lyon-Caen, H. *Spec. Pap. geol. Soc. Am.* **218**, 179-207 (1988).
12. Molnar, P. & Tapponnier, P. *J. geophys. Res.* **83**, 5361-5375 (1978).
13. England, P. C. & McKenzie, D. P. *Geophys. J. R. astr. Soc.* **70**, 295-321 (1982).
14. Houseman, G. A. & England, P. C. *J. geophys. Res.* **91**, 3651-3663 (1986).
15. Vilotte, J. P., Madariaga, R., Dagnières, M. & Zienkiewicz, O. *Geophys. J. R. astr. Soc.* **84**, 279-310 (1986).
16. England, P. C. & Houseman, G. A. *J. geophys. Res.* **91**, 3664-3674 (1986).
17. Peltzer, G. & Tapponnier, P. *J. geophys. Res.* (in the press).
18. Tapponnier, P., Peltzer, G., Le Dain, A. Y., Armijo, R. & Cobbold, P. *Geology* **10**, 611-616 (1982).
19. McKenzie, D. & Jackson, J. *Earth planet. Sci. Lett.* **65**, 182-202 (1983).
20. McKenzie, D. & Jackson, J. *J. geol. Soc. Lond.* **143**, 349-353 (1986).
21. Nur, A., Ron, H. & Scotti, O. *Geology* **14**, 746-749 (1986).
22. Burchfiel, B. C. *et al. Geol. Soc. Am. Bull.* (submitted).
23. Zhang Peizhen *et al. Geol. Soc. Am. Bull.* (submitted).
24. Beck, M. E. *Am. J. Sci.* **276**, 694-712 (1976).
25. Luyendyk, B. P., Kamerling, M. J., Terres, R. R. & Hornafius, J. S. *J. geophys. Res.* **90**, 12454-12466 (1985).
26. Ron, H., Freund, R., Garfunkel, Z. & Nur, A. *J. geophys. Res.* **89**, 6256-6270 (1984).
27. Kissel, C. & Laj, C. *Tectonophysics* **146**, 183-201 (1988).
28. Nelson, M. R. & Jones, C. H. *Tectonics* **6**, 13-33 (1987).
29. Nicholson, C., Seeber, L., Williams, P. & Sykes, L. R. *Tectonics* **5**, 629-648 (1986).
30. Sonder, L. J., England, P. C. & Houseman, G. A. *J. geophys. Res.* **91**, 4797-4810 (1986).
31. Lamb, S. H. *Earth planet. Sci. Lett.* **84**, 75-86 (1987).
32. Jackson, J. *Nature* **334**, 195-196 (1988).
33. Anderson, R. E. *Geol. Soc. Am. Bull.* **82**, 43-58 (1971).
34. Armstrong, R. L. *Geol. Soc. Am. Bull.* **83**, 1729-1754 (1972).
35. Wernicke, B. *Nature* **291**, 645-648 (1981).
36. Miller, E. L., Gans, P. B. & Garing, J. *Tectonics* **2**, 239-264 (1983).
37. Davis, G. A. & Lister, G. S. *Spec. Pap. geol. Soc. Am.* **218**, 133-159 (1988).
38. Jones, C. H. *Tectonics* **6**, 449-473 (1987).
39. Davis, G. A. & Burchfiel, B. C. *Geol. Soc. Am. Bull.* **84**, 1407-1422 (1973).
40. Cheadle, M. J. *et al. Tectonics* **5**, 293-320 (1986).
41. Lemiszki, P. J. & Brown, L. D. *Geol. Soc. Am. Bull.* **100**, 665-676 (1988).
42. Hadley, D. & Kanamori, H. *Geol. Soc. Am. Bull.* **88**, 1469-1478 (1977).
43. Humphreys, E., Clayton, R. W. & Hager, B. H. *Geophys. Res. Lett.* **11**, 625-627 (1984).
44. Panza, G. F. & Müller, St. *Mem. Sci. Geol.* **33**, 43-50 (1979).
45. Houseman, G. A., McKenzie, D. P. & Molnar, P. *J. geophys. Res.* **86**, 6115-6132 (1981).
46. Lyon-Caen, H. & Molnar, P. *J. geophys. Res.* **88**, 8171-8191 (1983).
47. Molnar, P. *Phil. Trans. R. Soc. A* **326**, 33-88 (1988).

ARTICLES

New semiconductor device physics in polymer diodes and transistors

J. H. Burroughes, C. A. Jones & R. H. Friend

Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

Semiconductor devices have been made from polyacetylene, a conjugated polymeric semiconductor. The device operates in a novel way: charge is stored in localized soliton-like excitations of the polymer chain, which are introduced not by doping or photoexcitation but by the presence of a surface electric field. The formation of charged solitons changes the optical properties of the polymer, introducing optical absorption below the band gap. Combined with the processibility of the polymer, these new electro-optic effects may be exploited technologically in electro-optic modulators.

THERE is current interest in the possible use of 'molecular' organic materials as the active materials in electronic or optical devices, and this field is now designated 'molecular electronics'. But basic questions as to whether or not it is possible (or appropriate) to make use of modes of operation familiar in inorganic semiconductor science remain largely unanswered. Conjugated polymers are 'molecular' analogues of inorganic semiconductors¹, which exhibit high electronic mobilities when doped². Despite the considerable progress made in the past decade towards an understanding of the electronic properties of conjugated polymers, there has been relatively little work on their use as the active component in semiconductor device structures. There are several reasons for this, the most important of which is that most conjugated polymers cannot be conveniently processed to the forms required in these devices. Most conjugated polymers are not readily soluble in easily handled solvents, and are infusible. This has severely limited the scope for construction of devices and those that have been reported in the literature show rather poor characteristics. Schottky and p-n diodes have been studied by several groups³⁻⁶, and rec-

tification ratios ($I_{\text{forward}}/I_{\text{reverse}}$) of up to a few hundred have been reported. There are also reports of field-induced conductivity measured in MISFET (metal-insulator-semiconductor field-effect transistor) structures^{7,8}.

A major advance in the control of the polymer processing is in the use of a solution-processible precursor polymer which can be converted to the conjugated polymer after processing. This was first demonstrated by Edwards and Feast⁹ for the preparation of polyacetylene (the Durham route) and it is polyacetylene produced by this route that we have used in the present studies. We have made thin-film devices by spin-coating the precursor polymer, poly((5,6-bis(trifluoro-methyl)-bicyclo-[2,2,2]octa-5,7-diene-2,3-diyl)-1,2-ethenediyl), in solution onto the required substrate, followed by heat treatment to convert to the polyacetylene by elimination of hexafluoroorthoxylene¹⁰⁻¹⁵. As we discuss later, the films of polyacetylene prepared in this way are p-doped (possibly with catalyst residues at chain ends) to a level suitable for device applications ($\sim 10^{16} \text{ cm}^{-3}$), and these dopants do not appear to be mobile under applied electric fields. We have not found it necessary, therefore, to dope the

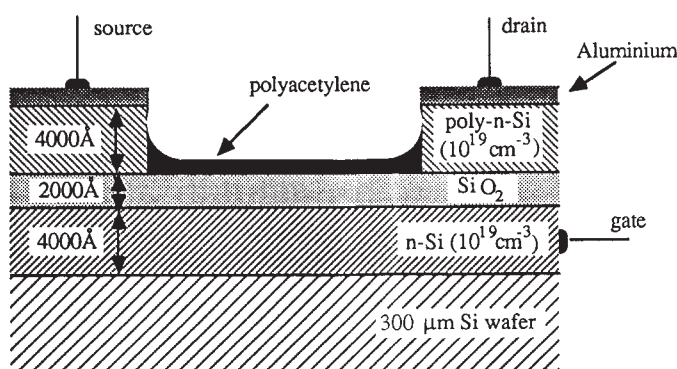


Fig. 1 Schematic diagram for a polyacetylene MISFET structure. Dimensions shown are to scale, except the channel width (20 μm) and length (1.5 m).

polymer further. We have investigated a range of unipolar devices: Schottky-barrier diodes, MIS (metal-insulator-semiconductor) structures and MISFETs. By careful control of the processing, with rigorous exclusion of oxygen, we have been able to get the device performance up to levels respectable enough to learn about the detailed functioning of the device. For the Schottky diodes we routinely measure rectification ratios of about 500,000, and for the MIS and MISFET structures we can demonstrate that the devices show 'textbook' formation of charge accumulation, inversion and depletion layers at the polyacetylene-insulator interface.

We did not expect charge storage in these devices to be in the band states as electrons and holes. It is established that there is a structural relaxation around charges present on the polymer chains, both when charges are added to the polymer chain through chemical doping, or separated following photoexcitation. For polyacetylene, the localized states are bond alternation defects, or solitons^{16,17}. Charge injected into the polyacetylene in these device structures should also be stored in soliton states. The formation of a depletion layer in polyacetylene should remove the charged solitons introduced through extrinsic doping. The formation of an accumulation or inversion layer, however, provides a new means of charging the polymer chains. The clearest evidence for the formation of solitons in polyacetylene is the appearance of additional optical absorption below the band gap, associated with the vibrational and electronic excitations of the solitons¹⁸. For polyacetylene prepared by the Durham route the 'mid-gap' absorption feature due to transitions between the 'mid-gap' levels of the solitons and the band edges is seen at ~ 1.0 eV for chemically doped samples¹³, and at 0.55 eV

for photoexcited charges¹⁵. We expect, therefore, that 'charge injection' will change the optical properties of the polyacetylene in these devices; depletion should reduce the mid-gap optical absorption whereas accumulation or inversion-layer formation should introduce mid-gap states onto previously undistorted chains, increasing the mid-gap optical absorption. We find both qualitative and quantitative agreement with these theoretical predictions.

Fabrication

Thin films of polyacetylene were formed by spin-coating films of the Durham precursor polymer from solution in 2-butanone. Transformation of the precursor polymer to the *trans* isomer of polyacetylene was achieved by heating the film to 80 $^{\circ}\text{C}$ for 12 h. By adjustment of the concentration of the precursor polymer solution and the conditions for spin-coating, fully dense, coherent films of polyacetylene with thicknesses in the range 200 $\text{\AA}$ to 1 μm were readily produced. All fabrication and measurements involving the polyacetylene were carried out with rigorous exclusion of oxygen. A variety of device structures were fabricated and examples are given below. Note that most of these are arranged to allow passage of sub-band-gap light through the layers.

Schottky diodes. The multilayer structure used was built up with a wide-field spectroil substrate, 200 $\text{\AA}$ gold, polyacetylene film (typically 1 μm), and a top contact of 200 $\text{\AA}$ aluminium. Gold forms an ohmic contact with p-type polyacetylene, aluminium the rectifying contact.

MIS diodes. Several configurations have been used. (1) Ultra-pure silicon substrate (thickness 0.3 mm, $<10^{13} \text{ cm}^{-3}$ charged impurities), 4,000 $\text{\AA}$ heavily phosphorus-doped ($>10^{19} \text{ cm}^{-3}$) silicon layer, 2,000 $\text{\AA}$ silicon dioxide (insulator layer), polyacetylene (typically 200–1,200 $\text{\AA}$) and a top contact of 200 $\text{\AA}$ gold. (2) Wide-field spectroil substrate, 200 $\text{\AA}$ gold, 1,500 $\text{\AA}$ poly(methyl-methacrylate) (insulator layer; spin-coated from dichloromethane solution), polyacetylene film (typically 200 $\text{\AA}$), 200 $\text{\AA}$ chromium and 100 $\text{\AA}$ gold.

MISFET structures. (1) Gold source and drain contacts: n-doped silicon substrate (distributed gate electrode), 2,000 $\text{\AA}$ silicon dioxide, gold source and drain electrodes (thickness 500 $\text{\AA}$, channel length 0.1 mm, channel width 75 mm), polyacetylene film (typically 200 $\text{\AA}$). Similar source and drain contacts can also be deposited by evaporation onto the MIS diode (2) above. (2) Poly n-silicon source and drain contacts: the structure used is shown as Fig. 1.

Electrical characterization

Schottky diodes. The current density versus voltage characteristics for Schottky diodes constructed as above are shown in Fig. 2, in which $\log |I|$ is plotted versus bias voltage for both forward and reverse biases. The ratio of forward to reverse current reaches a value of typically 5×10^5 at bias voltages of ~ 1.5 V. The usual parameterization for the variation of current density with bias voltage is $J = J_{\text{sat}} \exp(qV/nkT)$ for $V > 3kT/q$, where n is the ideality factor, which for a 'perfect' diode is equal to 1. As seen in Fig. 2, $n \approx 1.3$ at low forward biases. At higher forward biases the current is limited by the bulk resistance. In the thermionic emission model for conduction across the junction, the saturation current density, J_{sat} is given by $J_{\text{sat}} = A^* T^2 \exp\{q\psi/kT\}$, with A^* the effective Richardson constant ($120 \text{ A K}^{-2} \text{ cm}^{-2}$ for an effective electron mass of one), and ψ the barrier height¹⁹. By extrapolating from the linear region in Fig. 2, we estimate that $J_{\text{sat}} = 4.4 \times 10^{-13} \text{ A cm}^{-2}$, and hence $\psi = 1.28$ eV. The reverse current does not appear to saturate, contrary to the prediction of the thermionic emission model; this is probably caused by the formation of a small interfacial layer between the polyacetylene film and the blocking contact²⁰. The doping density may be obtained from the variation of the Schottky diode capacitance with reverse bias voltage, using the simple depletion model for the voltage dependence of the deple-

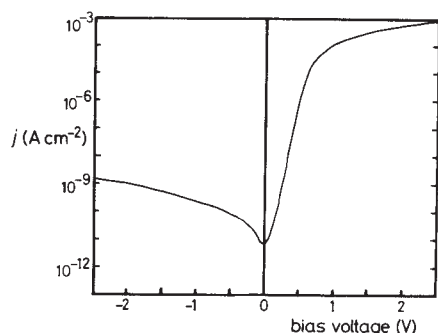


Fig. 2 Modulus of current versus voltage for a Schottky diode.

tion width. In terms of the measured differential capacitance, C , this may be expressed as

$$\frac{1}{C^2} = \frac{1}{A^2} \left(\frac{2(\phi + V)}{\epsilon_r \epsilon_0 q N_A} \right)$$

where N_A is the acceptor concentration and ϕ is the built-in potential¹⁹. This relation is obeyed well for reverse bias, giving values of $N_A = 9 \times 10^{15} \text{ cm}^{-3}$ and $\phi = 0.94 \text{ V}$. We thus find that the electrical characteristics of the Schottky barrier between aluminium and polyacetylene indicate that the device is operating in the conventional way, whereby the depletion layer boundary moves under the influence of the applied bias voltage. We show later, however, that the charge depleted in the polyacetylene is stored in soliton states, rather than as holes in the valence band, in which case detailed modelling of the electrical properties using a conventional, rigid-band picture is inappropriate.

MIS structures. The MIS structure allows the possibility of band bending at the insulator-semiconductor interface through the Fermi level to produce a surface-charge layer which may be of the same carrier sign as the majority carriers (accumulation layer) or as the minority carriers (inversion layer)²¹. The formation of accumulation, depletion and inversion layers may be demonstrated through the relationship between the device capacitance and the bias voltage. The measured capacitance, C , is that of the series combination of the insulator capacitance, C_i , and the capacitance of the active region of the semiconductor, C_d , and is given by $C = C_i C_d / (C_i + C_d)$. Because C_d is large for the accumulation and inversion layers, C is equal to the geometric capacitance of the insulating layer, but falls to a lower value if the depletion layer moves. The capacitance versus bias voltage curve for a polyacetylene MIS structure is shown in Fig. 3. As expected, for negative gate voltages the capacitance tends to the geometric capacitance of the insulator, indicating the formation of an accumulation layer. The decrease in measured capacitance for positive voltages reveals the formation of a depletion layer. Saturation of the capacitance sets in for large positive biases, when the depletion layer extends across the polyacetylene film (estimated to be $\sim 1,200 \text{ \AA}$ in thickness for this device).

MISFETs. A MISFET structure demonstrates that the charges at the polyacetylene-insulator interface in the MIS structure are mobile: source and drain contacts allow measurement of the conductance of the surface charge layer. Figure 4a shows results for a MISFET fabricated with gold source and drain contacts, for which the channel current, I_{DS} , is plotted against drain voltage, V_{DS} , for different gate voltages, V_{GS} . The channel current increases with negative gate voltage and saturates for large V_{DS} because of pinch-off of the channel near the drain electrode. Because the source and drain electrodes form ohmic

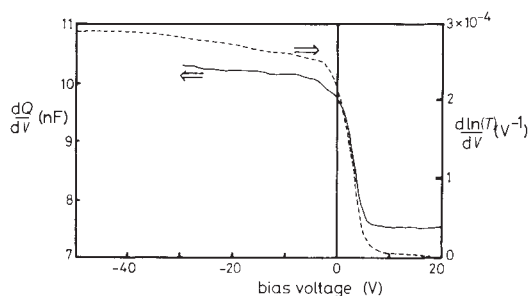


Fig. 3 The differential optical transmission, $\partial \ln(T)/\partial V$ at 0.8 eV (dashed), and differential capacitance, $\partial Q/\partial V$ (solid) versus bias voltage for the MIS structure. Measurements were made at 500 Hz and with an a.c. modulation of 0.25 eV .

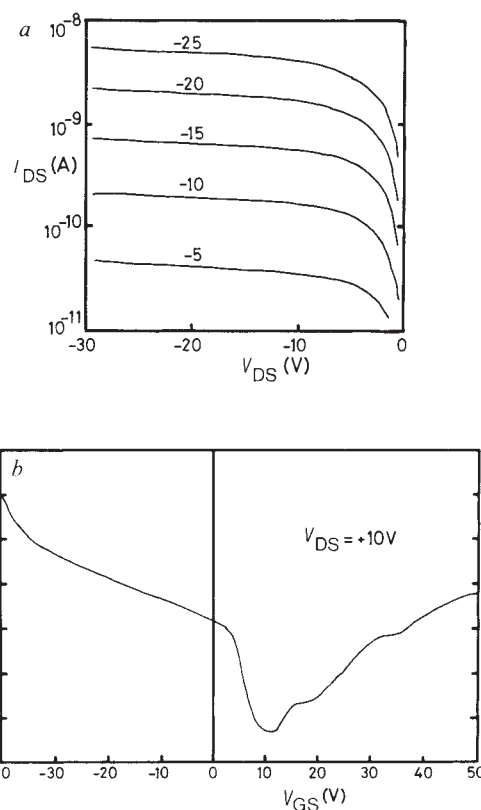


Fig. 4 *a*, I_{DS} versus V_{DS} at values of V_{GS} from -5 to -25 V , for a MISFET with gold source and drain contacts. *b*, I_{DS} versus V_{GS} at constant $V_{DS} (+10 \text{ V})$ for a MISFET with poly n-silicon source and drain contacts, as shown in Fig. 1.

contacts with the polyacetylene, we expect majority carrier injection from the electrodes into the polymer. The device is thus operating by modulation of the accumulation layer at the semiconductor-insulator interface. The channel conductivity for this structure could be modulated by a factor of about 2,000, from extreme depletion (full thickness of polyacetylene depleted) at $V_{GS} = +5 \text{ V}$ to accumulation at $V_{GS} = -30 \text{ V}$. Formation of an inversion layer for large positive gate voltages is not expected to increase the channel conductance since the gold source and drain contacts are expected to form blocking contacts to the n-polyacetylene layer.

The MISFET in Fig. 1 is constructed with n-silicon source and drain contacts, which are expected to make ohmic contacts to the n-polyacetylene layer. We thus expect to find enhanced channel conductance by formation of an n-type inversion layer for positive gate voltages. Figure 4b shows the variation of I_{DS} with V_{GS} at constant V_{DS} . The channel conductance has a minimum at $V_{GS} = +10 \text{ V}$ (full depletion) and rises for gate voltages both more positive (inversion layer) and more negative (accumulation layer), with a maximum on/off ratio for this structure of 100,000 ($V_{GS} = -40 \text{ V}$ to $+10 \text{ V}$). We see, therefore, the behaviour expected for positive bias, although the strong enhancement of the conductivity for negative bias indicates that the accumulation layer is still able to make good contact to the source and drain electrodes.

The MISFET structure provides a very convenient means of controlling the charge carrier concentration in the surface layer of the semiconductor. We have determined the carrier mobility from conductance measurements as a function of gate voltage, and find low values, typically $10^{-4} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. Similar mobilities are estimated for photogenerated carriers and for the extrinsic carriers in Durham polyacetylene (P. D. Townsend

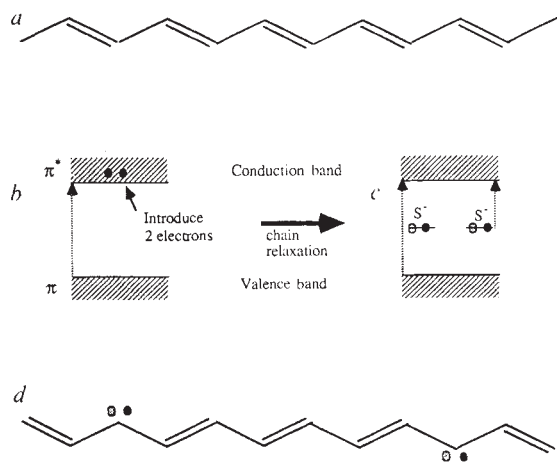


Fig. 5 Schematic representation of the formation of a pair of solitons on a polyacetylene chain. *a*, Undistorted chain; *b*, associated band scheme. The interband π - π^* optical transition is shown as a dotted arrow. If two electrons are introduced into the conduction band of the chain as shown in *b*, the chain relaxes to the form shown in *d*, with a reversed sense of bond alternation in the centre of the chain separated by two solitons. Associated with the two solitons are non-bonding π states in the gap, created from one (doubly occupied) valence and one (empty) conduction band state. These two states (S) are doubly occupied, as shown in *c*, and each carries a negative charge. New optical transitions from the soliton level to the conduction band are indicated with a dotted arrow; the oscillator strength for the interband transition is weakened through the loss of the band states. A complementary picture holds for positive charge, which is accommodated as unoccupied solitons levels.

and R. H. Friend, to be published); we consider that the mobility-limiting process is the transfer of charge between chains.

Optical properties

As discussed earlier, charge in polyacetylene is stored in 'soliton' localized states, which are associated with non-bonding p_z mid-gap states. As a result, we expect to see a modulation in the optical properties of the active semiconductor region of these devices with applied bias voltage. The formation of a pair of negatively-charged solitons on a polyacetylene chain is shown schematically in Fig. 5.

Schottky diodes. An increase in the width of the depletion layer should remove charged soliton states thereby reducing absorption within the gap and increasing the absorption in the region of the π - π^* transitions above 1.4 eV. The voltage-modulated transmission (VMT) spectrum between 0.4 and 1.8 eV is shown in Fig. 6. The positive signal for energies less than 1.4 eV shows that the transmission does increase for reverse bias voltages. The position of the electro-bleaching peak at 0.55 eV is in general agreement with photo-induced absorption experiments¹⁵ on spin-coated Durham polyacetylene. The periodic structure in the modulated transmission spectrum below the band gap is due to the formation of interference fringes (modulation of the finesse of the Fabry-Perot etalon formed by the polyacetylene film). Electro-absorption is seen at energies greater than 1.4 eV, as expected if extra π - and π^* -band states are created at the expense of soliton states.

MIS and MISFET structures. For these structures we expect a decrease in the device transmission below the band-edge as the device is driven towards accumulation and new charged solitons are introduced onto the polyacetylene chains at the interface with the insulator. The VMT spectrum between 0.4 and 1.2 eV (Fig. 7a) shows a decrease in the transmission

through the device for negative bias. The peak value of $\Delta T/T$, at 0.8 eV, is $\sim 0.64\%$. If all the charge at the polymer-insulator interface is stored in soliton-like states then the VMT signal, $\partial \ln(T)/\partial V$, should scale with $\partial Q/\partial V$, the differential capacitance, as the bias voltage is varied. This is shown to be the case in Fig. 3, where $\partial \ln(T)/\partial V$ and the differential capacitance are both plotted against bias voltage. We have investigated the VMT response in a variety of structures, including the MISFET shown in Fig. 1. For this structure we find results similar to those in Fig. 7a, that is, the VMT response expected for both accumulation and inversion. It should be noted that the optical path for the MISFET is between the source and drain contacts, and that there is no 'back' electrode.

For the low areal soliton concentrations, N^a , found in the device structures, the modulation in transmission, $\Delta T/T$, can be expressed as σN^a , where σ is the optical cross-section for the dopant-induced mid-gap absorption. The determination of N^a is particularly simple for the MIS device driven into accumulation, for which $N^a = \epsilon_r \epsilon_0 V/qd$, where V is the bias voltage and d the insulator thickness. The value for $\partial \ln(T)/\partial V$ becomes constant for negative bias voltages (Fig. 3), as expected in this simple model. Typical results, such as those shown in Fig. 7a, give the magnitude of the optical cross-section at the peak in the mid-gap absorption as $1.2 \times 10^{-15} \text{ cm}^2$, in very close agreement with the value measured from dopant-induced absorption¹⁸. We obtain similar agreement from analysis of optical modulation in the Schottky diode. Note that the maximum depth of optical modulation is fixed by the maximum value for N^a , which is determined for the MIS structure by the product $\epsilon_r E_{\text{max}}$ for the insulator, where E_{max} is the dielectric breakdown field. For silicon oxide, the maximum value for N^a is $\sim 2 \times 10^{13} \text{ cm}^{-2}$, and the maximum value for $\Delta T/T$ is thus $\sim 2.4\%$.

Besides the electronic transitions associated with the soliton state, there is extra infrared activity from new vibrational modes that couple to motion of the charged soliton along the polymer chain²². These are seen in both doping¹³ and photoexcitation¹⁵ experiments on unoriented Durham polyacetylene. Figure 7b shows the results from an FTIR experiment on an MIS structure, showing the difference in transmission for the MIS device biased between 0 and -50 V with respect to the gate. There are three peaks: an absorption above $4,000 \text{ cm}^{-1}$ (this is the low-energy side of the electronic absorption, as in Fig. 7a) and two sharp vibrational features at $1,379$ and $1,281 \text{ cm}^{-1}$. We identify these as the two higher-frequency translation modes of the soliton. The value of $1,379 \text{ cm}^{-1}$ lies between that measured for photo-induced charges, $1,373 \text{ cm}^{-1}$ (ref. 15), and for dopant-induced charges, $1,410 \text{ cm}^{-1}$ (ref. 13), and provides information about the chain conformation and degree of conjugation. We were not able to measure the VMT spectrum below $1,000 \text{ cm}^{-1}$ because

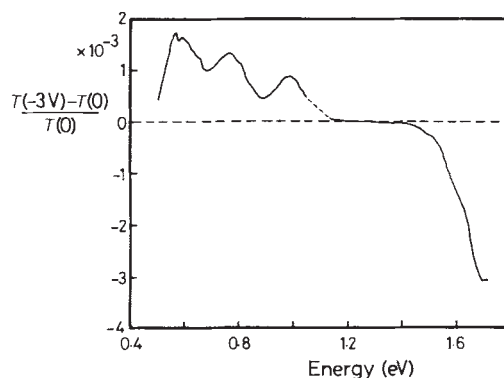


Fig. 6 Voltage-modulated transmission for a Schottky diode, $[T(V) - T(0)]/T(0)$, versus photon energy (E_p), with $V = -3 \text{ V}$ with respect to the blocking contact.

of a very large absorption peak from SiO_2 , and we were thus unable to observe the lowest-frequency, 'pinned' translational mode.

The results presented above are all obtained on structures with silicon dioxide as the insulator layer. It is important to demonstrate that these effects can be achieved with other insulators, and we have investigated several organic systems that can be readily formed as thin coherent films. Results on polyacetylene MIS structures, with spin-coated poly(methylmethacrylate) serving as the insulator, show similar behaviour, with an absorption band below the band gap, although the peak in absorption is at a lower energy, 0.55 eV.

Discussion

The formation of depletion, inversion and accumulation charge layers in a semiconductor device is restricted to semiconductors that are free of defects with energy levels within the semiconductor gap. Polymeric semiconductors are not obvious candidates since the structure is inherently disordered, with a large number of conformation defects (bends, twists) on the chains. But there are two principles common to conjugated polymers which run in their favour. First, there are no unsatisfied chemical bonds at the surface of the polymer, and we can expect that if the interface with the metal (Schottky diode) or insulator (MIS structures) is prepared in clean, oxygen-free conditions, we should avoid the problems with surface states well known for inorganic semiconductors. Second, defects on the polymer chain tend naturally to weaken the degree of π electron delocalization, and therefore to increase the π - π^* gap at the defect^{11,13,15}. The electronic levels associated with the defect are thus removed from the semiconductor gap and play no active part in the device operation. This principle is of particular importance here since the unoriented films of Durham polyacetylene we have used are known to have straight-chain sequences of no more than 20 or so carbon-carbon bonds¹⁰⁻¹⁵.

Optical spectroscopy reveals that charge storage in depletion, inversion or accumulation layers in these polyacetylene devices is not in band states (as with conventional semiconductors) but in soliton-like mid-gap states. For the depletion layer the chain deformation associated with these states is achieved by extrinsic doping. For accumulation and inversion layers, however, we have demonstrated a new means of putting charge onto the chains, in which the presence of the surface field forces charged soliton-antisoliton pairs onto previously undistorted chains, as shown schematically in Fig. 5.

Modulation of the optical properties by control of the charged-soliton density with the voltage applied across the device provides a new and powerful spectroscopy for investigating the electronic structure of the active region in these semiconductor devices. By comparison of the variation of differential transmission at mid-gap ($\partial \ln(T)/\partial V$) and the differential capacitance ($\partial Q/\partial V$) with bias voltage, we have shown that there is a direct correspondence between the two quantities and that therefore

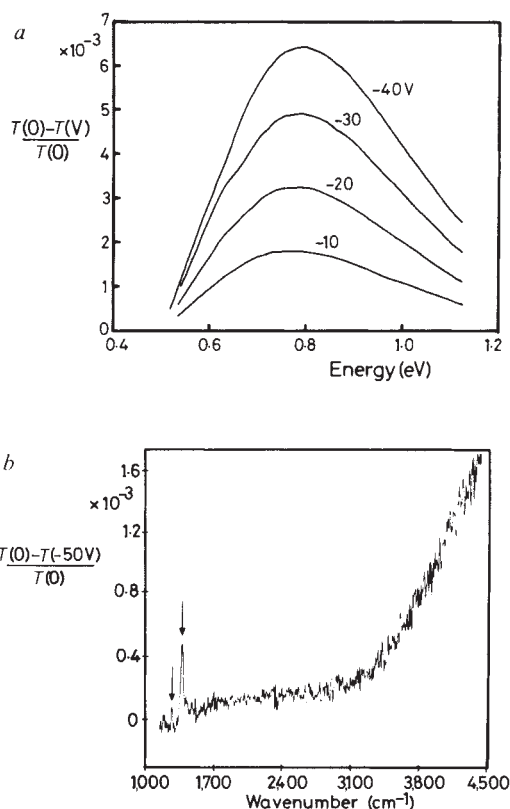


Fig. 7 *a*, Voltage-modulated transmission for a MIS diode, $[T(0) - T(V)]/T(0)$, versus photon energy, for various values of V (all negative with respect to the gate). *b*, Voltage-modulated transmission for a MIS diode, $[T(0) - T(-50V)]/T(0)$, in the spectral range 1,000–4,500 cm^{-1} .

all charges at the interface are accommodated on polyacetylene chains in soliton-like states. From the spectrally resolved electronic and vibrational absorption features, we can also learn much about the detailed electronic structure of those polymer chains at the interface on which the charge is stored; we will report on this elsewhere.

The electro-optic effects we have demonstrated here are large, and there are obvious areas of technology in which these devices may be exploited. Modulation depths, measured here for a single pass of the light beam through the device, can be readily enhanced, either in multiple-reflection structures (Fabry-Perot etalon) or in wave-guided structures in which the light is guided parallel to the active surface of the semiconductor.

We thank British Petroleum PLC for financial support for this work.

Received 6 July; accepted 10 August 1988.

1. *Handbook on Conducting Polymers* (ed. Skotheim, T. J.) (Dekker, New York, 1986).
2. Basescu, N. *et al. Nature* **327**, 403–405 (1987).
3. Grant, P. M., Tani, T., Gill, W. D., Krounbi, M. & Clarke, T. C. *J. appl. Phys.* **52**, 869–873 (1980).
4. Kanicki, J. in *Handbook on Conducting Polymers* (ed. Skotheim, T. J.) 544–660 (Dekker, New York, 1986).
5. Garnier, F. & Horowitz, G. *Synthetic Metals* **18**, 693–698 (1987).
6. Tomozawa, H., Braun, D., Phillips, S., Heeger, A. J. & Kroemer, H. *Synthetic Metals* **22**, 63–69 (1987).
7. Ebisawa, E., Kurokawa, T. & Nara, S. *J. appl. Phys.* **54**, 3255–3260 (1983).
8. Koezuka, H., Tsumura, A. & Ando, T. *Synthetic Metals* **18**, 699–704 (1987).
9. Edwards, J. H. & Feast, W. J. *Polymer Commun.* **21**, 595–597 (1980).
10. Friend, R. H. *et al. Trans. R. Soc. A314*, 37–49 (1985).

11. Friend, R. H. *et al. Synthetic Metals* **13**, 101–112 (1986).
12. Bott, D. C. *et al. Synthetic Metals* **14**, 245–269 (1986).
13. Friend, R. H., Bradley, D. D. C., Townsend, P. D. & Bott, D. C. *Synthetic Metals* **17**, 267–272 (1987).
14. Friend, R. H., Bradley, D. D. C. & Townsend, P. D. *J. Phys.* **D20**, 1367–1384 (1987).
15. Friend, R. H., Schaffer, H. E., Heeger, A. J. & Bott, D. C. *J. Phys.* **C20**, 6013–6023 (1987).
16. Su, W. P., Schrieffer, J. R. & Heeger, A. J. *Phys. Rev. Lett.* **42**, 1698–1701 (1979); **B22**, 2099–2111 (1980); erratum, **B28**, 1138 (1983).
17. Rice, M. J. *Phys. Lett.* **71A**, 152–154 (1979).
18. Orenstein, J. in *Handbook on Conducting Polymers* (ed. Skotheim, T. J.) 1297 (Dekker, New York, 1986).
19. Sze, S. M. *Physics of Semiconductor Devices*, 2nd edn (Wiley-Interscience, New York, 1981).
20. Rhoderick, E. M. *Metal-Semiconductor Contacts* (Clarendon Press, Oxford, 1978).
21. Ando, T., Fowler, A. B. & Stern, F. *Rev. mod. Phys.* **54**, 437–672 (1982).
22. Horovitz, B. *Solid St. Commun.* **41**, 729–734 (1982).

The origin of mutants

John Cairns, Julie Overbaugh & Stephan Miller

Department of Cancer Biology, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA

Nucleic acids are replicated with conspicuous fidelity. Infrequently, however, they undergo changes in sequence, and this process of change (mutation) generates the variability that allows evolution. As the result of studies of bacterial variation, it is now widely believed that mutations arise continuously and without any consideration for their utility. In this paper, we briefly review the source of this idea and then describe some experiments suggesting that cells may have mechanisms for choosing which mutations will occur.

WHEN populations of single cells are subject to certain forms of strong selection pressure, variants emerge bearing changes in DNA sequence that bring about an appropriate change in phenotype. If the selection is being applied to cells growing in liquid, the population with the original genotype will sooner or later be supplanted by the variants; if the cells are immobilized on a plate, the variants will form colonies or papillae, rising above the rest of the population. Our problem is to determine how many of these variants are arising as a direct and specific response to the selection pressure (would not have occurred in its absence) and how many are 'spontaneous' (would have arisen even in the absence of selection).

Luria and Delbrück were the first to attempt this distinction¹. They studied the origin of the phage-resistant mutants that are found when cultures of *Escherichia coli* are plated out in the presence of bacteriophage T1. They considered two extreme models. In the first model, the mutants are assumed to arise after plating, in response to the attack by the bacteriophage; the mutations occur in a small fraction of the bacteria exposed to the virus, and the number of mutants found in each of a large group of independent cultures should form a Poisson distribution whose mean is simply the product of the number of bacteria plated and the probability that a bacterium can become resistant before it is killed by the virus. In the second model, the mutants are assumed to arise spontaneously, as the result of a rare error in replication that has a constant probability of occurring in the life of each bacterium; in this case, the mutational events will be distributed randomly throughout each culture's previous history (scattered at random through its family tree), and the final number of mutants should be highly variable because occasional cultures will, by chance, have suffered a mutation early in their growth and will therefore end up containing a very large number of mutants. When put to the test, the number of T1-resistant mutants was observed to vary greatly from culture to culture, with the occasional culture containing up to 100 times the median value for the group as a whole. Luria and Delbrück concluded that mutation to phage resistance is the result of rare spontaneous events, occurring during the prior growth of cultures before they had been exposed to any selection.

By various ingenious sampling procedures such as 'replica plating', which tested cells' siblings rather than the cells themselves^{2,3}, it subsequently became possible to prepare pure cultures of mutants (for example, streptomycin-resistance mutants) without at any time having exposed the cells or their ancestors to any selection pressure. These experiments constituted the final proof that at least some forms of bacterial mutation are occurring spontaneously, before the bacteria can have any indication of each mutation's possible utility. They had, however, shown only that some mutants do arise spontaneously, in the absence of selection. In fact, we now know that mutations to phage and streptomycin resistance are not expressed until several generations after the change in DNA sequence has occurred⁴; thus the resistant mutants one isolates by spreading

cultures on selective plates are the result of events that must have occurred many generations before one applies the selection that demonstrates their existence. So these classical experiments could not have detected (and certainly did not exclude) the existence of a non-random, possibly product-oriented form of mutation.

To look for such a process, we have studied mutations that are immediately expressed, and where the selection pressure rewards mutants by letting them multiply but allows all the other, non-mutant cells to survive so that they can at least have the opportunity to perform directed mutation. For this analysis, it was necessary to calculate the expected composite distribution, where some mutants are arising spontaneously (during the prior growth of each culture) and some are arising later, after the cells are put on selective plates.

Mutant distribution in independent cultures

We have to consider two distributions and their combination. The first distribution is created when mutants arise at random while the population in each culture is increasing by binary fission; at this stage, before any selection has been applied, mutants and non-mutants are postulated to multiply at the same rate; when the cultures are put out on selective plates, the mutants are able to form colonies. In these circumstances, the distribution of mutants among a set of independent cultures will be determined by a single parameter, the constant probability of mutation per cell per generation. Lea and Coulson⁵ were the first to derive an exact solution. The distribution has certain awkward properties^{6,7} but the following argument leads to a useful approximation for its upper end. For a culture undergoing binary fission, the number of cells (n) present at any particular moment plus the number present in all previous generations (that is, n plus $n/2 + n/4 + n/8 + \dots$) is twice as great as the number present in all previous generations ($n/2 + n/4 + n/8 + \dots$); so the chance that any particular mutation occurred at least $g+1$ generations ago is twice the chance that it occurred at least g generations ago. If, therefore, we just consider those very rare cultures that suffered a mutation so early in their history that in each culture one single clone (that is, one single mutational event) accounts for virtually all the mutants finally present, we see that the proportion of such cultures that contain at least 2^g cells will be twice the proportion with at least 2^{g+1} ; this describes exactly the upper end of the distribution. In fact, if the cultures finally contain N cells and have suffered an average of m mutational events per culture (the mutation rate is $m/2N$ per cell), the probability (P_j) of finding a 'jackpot' of at least j mutants approaches m/j . This approximation had previously been derived for the case where m is very small⁸, but actually it holds for the upper end of the distribution irrespective of the value of m . Figure 1a shows the exact relation between $\log x$ and $\log P_x$ (the logarithm of the proportion of cultures with x or more mutants) for several values of m .

A very different distribution will be created if, instead, all the

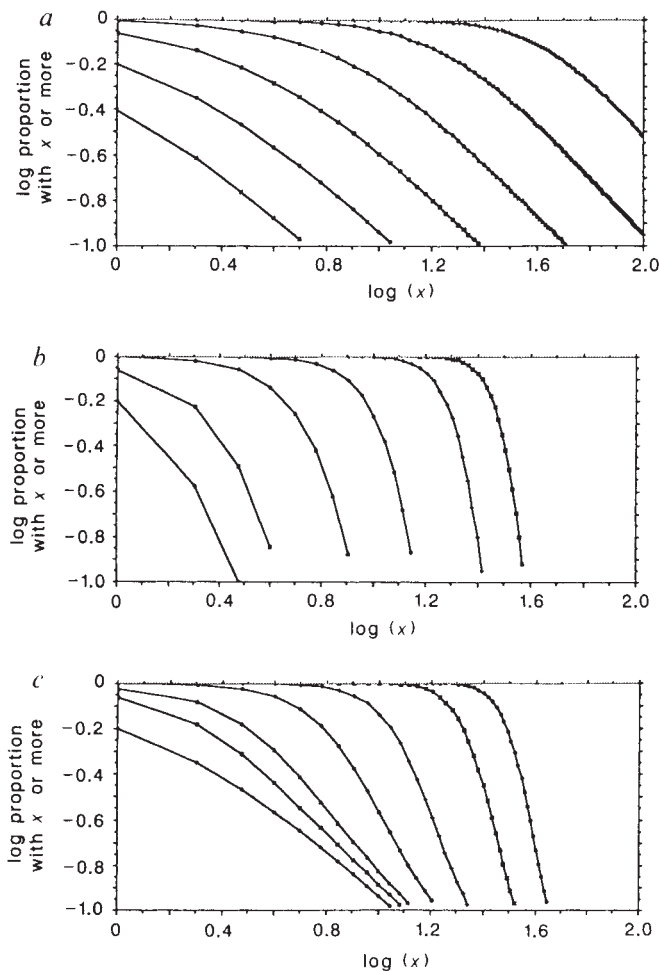


Fig. 1 The expected distributions of mutants among sister cultures that have experienced various kinds of mutational process. Each shows how the logarithm of the probability that a culture has at least x mutants ($\log_{10} P_x$) declines as $\log_{10} x$ increases. *a*, The case where the mutational events are distributed randomly among the entire lineage of each culture (calculated from Lea and Coulson Eq-15, ref. 5). The six examples shown are where the mean number of events per culture (m) is 0.5, 1, 2, 4, 8 and 16. *b*, The case where the mutational events are distributed randomly among the last generation of cells (that is they form a Poisson distribution). The six examples shown are where the mean number of events per culture (μ) is 1, 2, 5, 10, 20 and 30. *c*, Some combinations of the two previous distributions (where m events occur at random throughout the previous lineage of each culture and μ additional events occur in the final generation, among the cells that are being tested for the presence of the mutation). The examples show the expected distribution when $m = 1$ and $\mu = 0, 1, 2, 5, 10, 20$ and 30.

mutants arise after plating, among the bacteria exposed to the selection pressure. The numbers in independent cultures should form a Poisson distribution, and some examples are shown, for comparison, in Fig. 1*b*. The distributions are much narrower and $\log P_x$ falls rapidly as $\log x$ increases, especially when the average number of mutations per culture is high.

If both forms of mutation are occurring, the distribution of $\log P_x$ will be a composite of the two previous distributions. The lower end of the distribution will tend to be dominated by the mutations arising after selection, so that $\log P_x$ will initially fall steeply with increasing x ; at high values, the distribution will be dominated by the rare jackpots produced by the mutations that occurred early during prior growth, and P_x will be inversely proportional to x . Various composite distributions are shown in Fig. 1*c*.

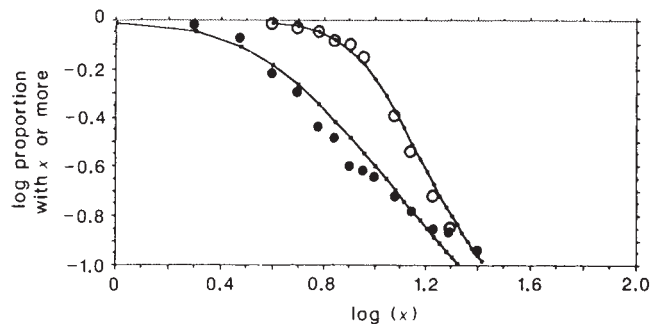


Fig. 2 The observed distribution of Lac⁺ revertants for *E. coli* $\Delta lac pro F' lacZ_{am}$ (filled circles) and for a derivative bearing a deletion of *uvrB* and *bio* (open circles). Sixty 3-ml cultures of each strain, in M9 medium plus 0.05% yeast extract, were grown overnight with aeration at 37 °C, and then each culture was centrifuged, resuspended in M9 and plated on M9-lactose plates. The colonies were counted after 48 h. The curves show the expected distributions for $m = 1.6$, $\mu = 2$ (on the left), and $m = 1.6$, $\mu = 7$ (on the right). (Roughly a third of the revertants, produced by these strains, are amber suppressors, and these form slightly smaller colonies at 48 h.)

Mutations affecting lactose fermentation

Bacterial variation in colony morphology was observed soon after the invention of plating procedures. One of the earliest studied examples was the formation of papillae of lactose-fermenting mutants, seen when non-fermenting strains of *E. coli* are plated out on rich media containing lactose, and it was attributed simply to a process of random variation and survival of the fittest⁹. This is a mutation that presumably is expressed as soon as it occurs, and so it is suitable for our present purposes.

Following the publication of tables for the Luria-Delbrück distribution⁵, Ryan studied the distribution of Lac⁺ variants in replicate cultures of Lac⁻ strains spread on lactose-minimal plates and, surprisingly, he found that in some strains it was narrower than expected¹⁰. We have confirmed this. The shape of the distribution of Lac⁺ revertants of strains bearing an amber mutation in *lacZ* depends on the genotype of the strain and on the medium in which it was growing before being plated; in certain circumstances, the distribution looks rather like some of those shown in Fig. 1*c*, as if there are two periods when mutations are occurring—first during the period of growth (producing the kind of distribution shown in Fig. 1*a*) and then in stationary phase (producing the kind of distribution shown in Fig. 1*b*). It is the second of these that is determined by the medium and genotype. In separate experiments, not described here, we have shown that such stationary phase mutation seems to be under the control of genes in the *uvrB-bio* region of the chromosome, but we have so far not been able to derive any simple picture of the regulatory process.

When multiple cultures of *E. coli lacZ_{am}* and of the same strain bearing a deletion in the *uvrB-bio* region are grown in yeast extract or in M9-glycerol and then plated on minimal-lactose plates, the two strains appear to have the same mutation rate during growth (they are indistinguishable at the upper ends of their distributions), but the *uvrB* strain adds a large number of extra mutants that are apparently part of a Poisson distribution (Fig. 2). Interestingly, these extra mutants tend to be slower to produce colonies than the mutants that are members of jackpots (that is, slower than the mutants already in existence at the time of plating). This suggests that the extra mutant colonies are the result of late events, occurring on the plates when the bacteria are in the presence of lactose and under strong selection pressure to become Lac⁺. The *uvrB* strain produces, each day about 3–5 times as many late mutants as the *uvr⁺* strain and it is therefore the strain of choice, when testing for the possible existence of directed mutation.

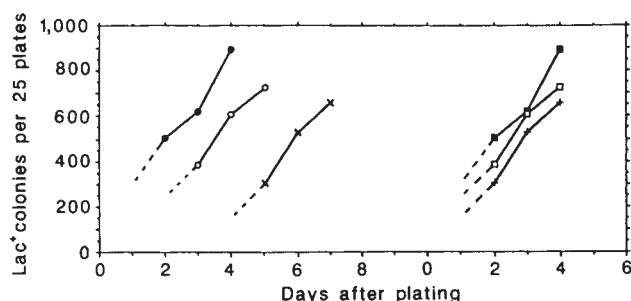


Fig. 3 The accumulation of Lac⁺ revertants of *E. coli* Δ *uvrB*-*bio* Δ *lac pro F' lacZ_{am}* after exposure to lactose on M9 plates. Six large cultures were grown overnight in M9 plus 0.1% glycerol. Fifteen 1-ml aliquots from each culture were plated with 2.5 ml top-layer agar on to M9-agar plates; the 90 plates were then overlaid with top-layer agar to ensure that all bacteria were buried in the agar; five plates from each culture were overlaid immediately with top-layer agar containing lactose, five were overlaid at 24 h, and five at 72 h. One of the six cultures proved to have a small jackpot of mutants, and so its plates were discarded; the plates from the remaining five cultures were counted each day, and the total counts are shown in real time on the left and in relation to time after the addition of lactose on the right.

To determine the role of lactose in the production of late Lac⁺ mutants, aliquots of several replicate cultures of the *uvrB* strain were plated on M9 plates without lactose. When the plates were overlaid immediately with top-layer agar containing lactose, they showed an average of 20 colonies by 48 h and then gained an extra 8 colonies a day thereafter. If, however, the addition of lactose was delayed one or three days, the whole time course of appearance of colonies was delayed one or three days; this delay was seen even when the plates contained the non-hydrolysable inducer of the *lac* operon, isopropyl-thiogalactoside. In short, the accumulation of late Lac⁺ mutants occurred only in the presence of lactose (Fig. 3).

One trivial explanation could be that this Lac⁺ strain can grow slowly in lactose (either because its amber mutation in *lacZ* is leaky or because the lactose is impure), and that is why it can continue to produce mutants on plates containing lactose. We have therefore tested for the presence of another rapidly expressed mutation, valine-resistance (Val^R), by overlaying the lactose plates at various times with agar containing glucose and valine, and we find that the number of Val^R colonies countable two days after overlaying with valine actually goes slightly downwards with time (just as did the number of Lac⁺ colonies in the absence of lactose). In other words, a population of bacteria that is accumulating Lac⁺ revertants on a lactose-minimal plate is not, at the same time, accumulating Val^R mutants.

This experiment suggests that populations of bacteria, in stationary phase, have some way of producing (or selectively retaining) only the most appropriate mutations. For *E. coli*, under pressure to revert one particular amber codon, 'appropriate mutation' does not add much to the high background mutation rate arising from errors during replication. The next two examples, however, suggest that it plays a much larger role when the selection pressure is for complex changes in genotype that must seldom if ever arise from chance errors in replication.

The Shapiro deletion

In order to study the process of spontaneous deletion, Shapiro¹¹ constructed a strain in which *araC* (the positive regulatory element controlling the arabinose operon) is placed upstream of *lacZ* but separated from it by a short segment of Mu bacteriophage DNA that contains transcription terminating signals. Both the arabinose and the lactose operons are missing in this

strain, so it is Lac⁻ and Ara⁻; if, however, it can delete the intervening Mu segment it will be able to grow on lactose, provided arabinose is present—a phenotype we will refer to as Lac(Ara)⁺. Shapiro has shown that the deletion occurs in populations of cells plated out on lactose-arabinose-minimal plates. The Lac(Ara)⁺ mutants are not, however, the result of spontaneous deletions occurring during the prior growth of the cultures but are apparently arising on the plates, because all the colonies appear much later than the colonies that are formed when cultures are seeded with a few pre-existing Lac(Ara)⁺ cells and then grown up and plated. Indeed, the frequency of spontaneous deletion during growth is probably less than 10⁻¹¹, because immediate-colony-formers have never been found in cultures grown in the absence of selection.

Because this appears to be potentially a case of directed mutation, we have investigated it a little further. The delay in the formation of Lac(Ara)⁺ colonies on selective plates is not simply some problem, in stationary phase, with the expression of *araC* and *lacZ* following their fusion, for we have prepared fused strains that are Lac(Ara)⁻ because they have point mutations in *lacZ* or *araC*, and these strains show no such delay in the appearance of their Lac(Ara)⁺ revertant colonies. Thus the act of gene fusion is indeed a late event, occurring long after the cells are put on lactose-arabinose-minimal plates. The event is influenced by some special property of the segment of Mu DNA that is being deleted, because the deletion does not occur when there is Mu DNA elsewhere in the chromosome¹¹. Even so, we would like to know if the deletion is actually dependent on the presence of lactose and arabinose.

We have therefore studied the appearance of Lac(Ara)⁺ cells (cells that are immediate-colony-formers on lactose-arabinose-minimal plates) in liquid cultures that are kept gently aerated after reaching stationary phase. The results can be summarized as follows. (1) In rich media containing both lactose and arabinose, large numbers of Lac(Ara)⁺ cells start to appear after 3–4 days at 30 °C. (2) Under these conditions, even minority populations (marked with rifamycin-resistance) contribute to the final population of Lac(Ara)⁺ cells, and from this we have calculated that the proportion of bacteria that are becoming Lac(Ara)⁺ must be at least 10⁻⁸. (3) In rich media without lactose (or without arabinose), Lac(Ara)⁺ cells do not accumulate at a detectable rate. (4) When lactose is added to a stationary culture that has arabinose but not lactose, Lac(Ara)⁺ cells promptly start to accumulate; judging from their rate of increase, the first Lac(Ara)⁺ cell must be formed within 1–2 h of the addition of lactose.

This therefore seems to be another example of the production of appropriate mutations in response to selection. In one sense, it is a rather clear case, because this particular mutation does not occur at a measurable rate in the absence of selection; so bacteria, in stationary phase, do apparently have access to some process that either can prevent useless mutations from occurring or can destroy unsuccessful mutants soon after they arise. At the same time, it is perhaps a rather special case, because it involves the movement (deletion) of a piece of Mu DNA and does not occur if there is Mu elsewhere in the chromosome.

So far we have only considered examples of spontaneous mutation where bacteria are under pressure to reverse defects that have been previously introduced into their genes by mutagenesis or by genetic engineering. For our final example, we will briefly consider the response of *E. coli* to more natural forms of selection pressure.

Activation of the cryptic genes in *E. coli*

The usual way of distinguishing the various pathogenic and nonpathogenic *Enterobacteriaceae* is to observe which sugars they can use as a source of energy; for example, *E. coli* ferments lactose, whereas the various members of the *Shigella* and *Salmonella* families do not. The standard test is to inoculate a small volume of peptone plus sugar, and observe turbidity, pH and

CO₂ production after 1–2 days. When this system of classification was being prepared, early this century, certain bacterial species were observed to be 'late' fermenters of some sugars. Thus it is a regular feature of some bacteria that it takes a week or more before the population of cells, produced by growth on the peptone, are able to generate a variant that can hydrolyse the sugar; for example, *Sh. sonnei* is classified as a 'late' fermenter of lactose, and many strains of *E. coli* are 'late' fermenters of salicin (an aromatic β -glucoside). The genes coding for the appropriate enzymes are there, but they are not readily accessible.

Bacteria apparently have an extensive armoury of such 'cryptic' genes that can be called upon for the metabolism of unusual substrates. The mechanism of activation varies. In some cases it is simply by the movement of an insertion sequence into a position upstream of the cryptic gene¹², but in others it may require several changes in base sequence. It is these latter examples that are of special interest. For instance, *E. coli* turns out to have a cryptic gene (*ebgA*⁰) that it can call upon to hydrolyse lactose, if the usual gene for this purpose (*lacZ*) has been deleted¹³. The activation of *ebg* requires at least two mutations, one in the repressor (*ebgR*) and one at a particular site in the gene coding for the enzyme (*ebgA*) to make the enzyme capable of hydrolysing lactose¹⁴. During growth, each of these point mutations occurs at a frequency of less than 10⁻⁸; neither on its own will allow a *lacZ* deletion strain to use lactose¹⁵, yet colonies of such a strain, that have grown to stationary phase on a plate containing lactose, will produce Lac⁺ papillae in about two weeks. That such events ever occur seems almost unbelievable, but we have also to realize that what we are seeing probably gives us only a minimum estimate of the efficiency of the process, since in these cases the stimulus for change must fairly quickly disappear once a few mutant clones have been formed that can exploit the novel sugar. It is difficult to imagine how bacteria are able to solve complex problems like these—and do so without, at the same time, accumulating a large number of neutral and deleterious mutations—unless they have access to some reversible process of trial and error.

Discussion

The main purpose of this paper is to show how insecure is our belief in the spontaneity (randomness) of most mutations. It seems to be a doctrine that has never been properly put to the test. We describe here a few experiments and some circumstantial evidence suggesting that bacteria can choose which mutations they should produce. But we realize that this is too important an issue to be settled by three or four rather ambiguous experiments.

The origin of genetic variation has been the subject of bitter controversy, throughout the nineteenth century and into the first half of the twentieth. Is all variation essentially random, like thermal noise? Or can the genome of an individual cell profit by experience? At its extremes, it was an argument between reductionists and romantics—between those who sought to explain the evolution and behaviour of the biosphere in terms of the laws of physics, and those who wished to make the success

of evolution just another manifestation of the mysteriousness of living things.

The early triumphs of molecular biology strongly supported the reductionists. The DNA double helix is an extraordinarily stable structure, but it is subject to spontaneous degradation and to errors in replication, and these changes are, in a sense, due to thermal noise. Furthermore, the discovery of all the elements that lie between DNA sequence and protein structure gave rise to the central dogma of molecular biology, and this doctrine denies any possible effect of a cell's experience upon the sequence of bases in its DNA.

Curiously, when we come to consider what mechanism might be the basis for the forms of mutation described in this paper we find that molecular biology has, in the interim, deserted the reductionist¹⁶. Now, almost anything seems possible. In certain systems, information freely flows back from RNA into DNA; genomic instability can be switched on under conditions of stress, and switched off when the stress is over; and instances exist where cells are able to generate extreme variability in localized regions of their genome. The only major category of informational transfer that has not been described is between proteins and the messenger RNA (mRNA) molecules that made them. If a cell discovered how to make that connection, it might be able to exercise some choice over which mutations to accept and which to reject.

Since this is the kind of versatility and adaptability we seem to be seeing in these experiments with *E. coli*, it is worth considering briefly how such a connection might be made. In a very direct way, the cell could produce a highly variable set of mRNA molecules and then reverse-transcribe the one that made the best protein; for this, it would have to have a special organelle (perhaps like the gag-pol complex of retroviruses) which contains reverse transcriptase plus some element that somehow monitors the protein product and determines whether the mRNA should go on being translated or should be transcribed into DNA; this could be an efficient process and makes an attractive hypothesis because it gives us an unusual explanation for the origin of retroviruses. Less efficient would be the production of reverse transcripts at random, but even that could achieve the desired end result, if reverse transcription were switched on temporarily, just at the time the cell resumed growth; any cell that started growing would acquire a reverse transcript of the variant sequence that made it able to grow, and so the sequence would eventually be incorporated by recombination into the genome of one of the cell's descendants (indeed, the same result could perhaps be achieved starting with DNA copies of segments of the genome rather than RNA transcripts). Each of these processes would allow individual cells to subject a subset of their informational macromolecules to the forces of natural selection. Each could, in effect, provide a mechanism for the inheritance of acquired characteristics.

We thank Victoria Cairns and Rainer Riester for writing a program for calculating the composite distributions of mutants, Patricia Foster and Richard Peto for comments on the manuscript, and Eric Eisenstadt for the hospitality of his laboratory. The work was supported by the National Science Foundation.

Received 2 May; accepted 11 July 1988.

1. Luria, S. E. & Delbrück, M. *Genetics* **28**, 491–511 (1943).
2. Lederberg, J. & Lederberg, E. M. *J. Bact.* **63**, 399–406 (1952).
3. Cavalli-Sforza, L. L. & Lederberg, J. *Genetics* **41**, 367–381 (1956).
4. Hayes, W. *The Genetics of Bacteria* (Blackwell, Oxford, 1964).
5. Lea, D. E. & Coulson, C. A. *J. Genet.* **49**, 264–285 (1949).
6. Armitage, P. *J. R. statist. Soc. B* **14**, 1–40 (1952).
7. Mandelbrot, B. *J. appl. Prob.* **11**, 437–444 (1974).
8. Luria, S. E. *Cold Spring Harb. Symp. quant. Biol.* **16**, 463–470 (1951).

9. Hadley, P. *J. infect. Dis.* **40**, 1–312 (1927).
10. Ryan, F. J. *Nature* **169**, 882–883 (1952).
11. Shapiro, J. A. *Molec. Gen. Genet.* **194**, 79–90 (1984).
12. Reynolds, A. E., Felton, J. & Wright, A. *Nature* **293**, 625–629 (1981).
13. Campbell, J. H., Lengyel, J. A. & Langridge, J. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1841–1845 (1973).
14. Hall, B. G. *J. Bact.* **129**, 540–543 (1977).
15. Hall, B. G. & Zuzel, T. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3529–3533 (1980).
16. Reanney, D. C. *Nature* **307**, 318–319 (1984).

Disruption of galactic radio jets by shocks in the ambient medium

Michael L. Norman^{*†}, Jack O. Burns^{*‡}
& Martin E. Sulkanen^{*‡}

^{*} National Center for Supercomputing Applications,
University of Illinois at Urbana-Champaign, Urbana,
Illinois 61801, USA

[†] Department of Astronomy, University of Illinois
at Urbana-Champaign, Urbana, Illinois 61801, USA

[‡] Institute for Astrophysics, Department of Physics and Astronomy,
University of New Mexico, Albuquerque, New Mexico 87131, USA

Observations show that jets in moderate luminosity ($<10^{25}$ W Hz⁻¹ at 20 cm wavelength) radio galaxies can flare dramatically in a few jet diameters and with opening angles up to 90° into diffuse lobes or tails¹⁻⁴ (Fig. 1). These morphologies are difficult to reproduce in numerical simulations of supersonic jets that move outward in constant^{5,6} or smoothly varying atmospheres (N. Norman, unpublished results). By analogy with structures seen in the laboratory⁷, one could interpret the collimated jets as moderately supersonic (Mach number 2-5) fluid flows, and the lobes or tails as subsonic plumes that are subject to turbulent broadening and entrainment of the ambient medium. Our studies indicate that the transition from supersonic to subsonic flow can occur suddenly only if a strong planar (normal to the flow direction) shock or Mach disk forms within the jet. Here we show that such an internal shock is produced as the jet crosses a shock wave in the external medium. The external shock could form, for example, by a galactic wind encountering the intergalactic medium. We find that for jet disruption the jet Mach number must be less than the wind-shock Mach number, a result also understandable from analytic arguments. We apply this model to Cen A and wide-angle-tailed radio galaxies.

Figure 2 shows schematically our numerical simulation. A rectangular domain contains initially two constant states of gas connected by a single, left-facing shock wave of strength (pressure ratio) P_2/P_1 . A unique choice of boundary values at the left- and right-hand sides of the computational box, which satisfy the Rankine-Hugoniot jump conditions, maintain the shock in a steady state. At time $t=0$, a supersonic, perfectly collimated jet is continuously injected through a central region of the left-hand boundary. Because the flow is taken to be inviscid and adiabatic, the problem is completely specified by the dimensionless ratios given in the Fig. 2 legend.

The jet at four stages of development is shown in Fig. 3. In Fig. 3a, the jet head is just crossing the external shock front. The ram pressure of the supersonic wind, blowing from left to right, has blown the jet cocoon forward into a lobe with strong internal eddies. By Fig. 3b, the essential structures of the shock-disrupted jet have formed. The external shock wave has been transmitted into the jet channel and is perpendicular to the flow direction. As a result, the post-shock jet flow is subsonic.

The jet disruption and flaring can be understood from simple analytical arguments in two-dimensional gas dynamics. First, the pressure jump across a shock depends only upon the upstream Mach number, $P_2/P_1 = 2\gamma M_1^2(\gamma+1)/(\gamma+1)$. If the Mach number of the wind shock, $M_{wind,1}$, exceeds that of the jet, $M_{jet,1}$, the post-shock jet pressure will be less than the post-shock wind pressure. By the Bernoulli theorem, the post-shock jet must further decelerate to increase its pressure to match that of the external medium. A subsonic flow expands as it decelerates, unlike a supersonic flow. This is seen in Fig. 3b-d.

Numerical simulations covering a range of $M_{jet,1}$ and $M_{wind,1}$ confirm that the condition for jet disruption is $M_{jet,1}/M_{wind,1} < 1$.

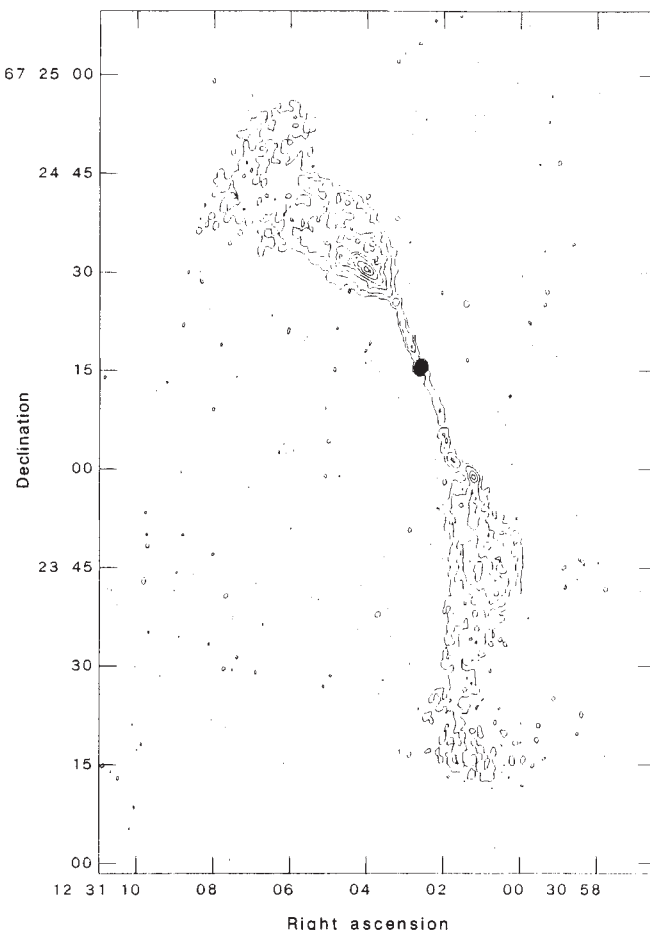


Fig. 1 VLA map of the WAT radio source 1231 ± 674, associated with the central dominant galaxy in Abell cluster 1559 (ref. 4). Clean beam is 1.15" × 0.93" at position angle -15°. The jets flare suddenly at ~10" from the compact core. (Map courtesy of A. O'Donoghue.)

This does not imply a velocity ratio $V_{jet,1}/V_{wind} < 1$, provided the sound speeds are such that $C_{jet,1} > C_{wind,1}$, which is generally thought to be the case in radio jets⁸. For larger values of this ratio, the jet remains collimated and supersonic, and terminates at a working surface or hot spot forming a classical double morphology (if large-scale fluid instabilities are absent). Jets in more powerful sources, such as Cyg-A, are thus expected to pass through an external shock unaffected by local galaxy conditions.

Second, the external shock near the jet must become oblique to divert the wind around the expanded subsonic jet. Referring to Fig. 2, the oblique angle of the shock, ϕ , can be found by equating the pressure change through the oblique shock to that of the jet shock; $\sin \phi = M_{jet,1}/M_{wind,1}$. The angle of the contact discontinuity, ψ , separating jet and ambient material is given by⁹

$$\cot \psi = \frac{M_{jet,1}}{(M_{wind,1}^2 - M_{jet,1}^2)^{1/2}} \left(\frac{(\gamma+1)M_{wind,1}^2}{2(M_{jet,1}^2 - 1)} - 1 \right) \quad (1)$$

The predicted angles $\phi = 60^\circ$ and $\psi = 27^\circ$, for the jet parameters of Fig. 2, agree well with those seen in Fig. 3, $\phi = 55 \pm 5^\circ$ and $\psi = 26 \pm 5^\circ$.

Third, the above analysis assumes the flow is laminar, but the velocity vectors in Fig. 3 show that turbulent eddies form, as expected, in the post-shock jet. An animation of the evolution shows extensive entrainment of external gas into the turbulent tail, although we have not yet quantified this phenomenon. Such turbulence is expected to expand the jet, especially as the ratio of internal to external densities, η , and the jet velocity decreases.

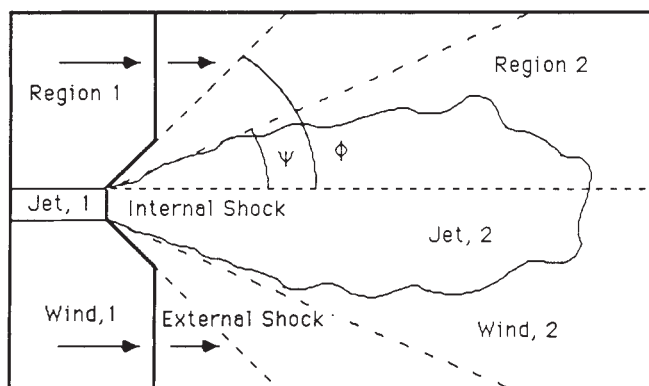
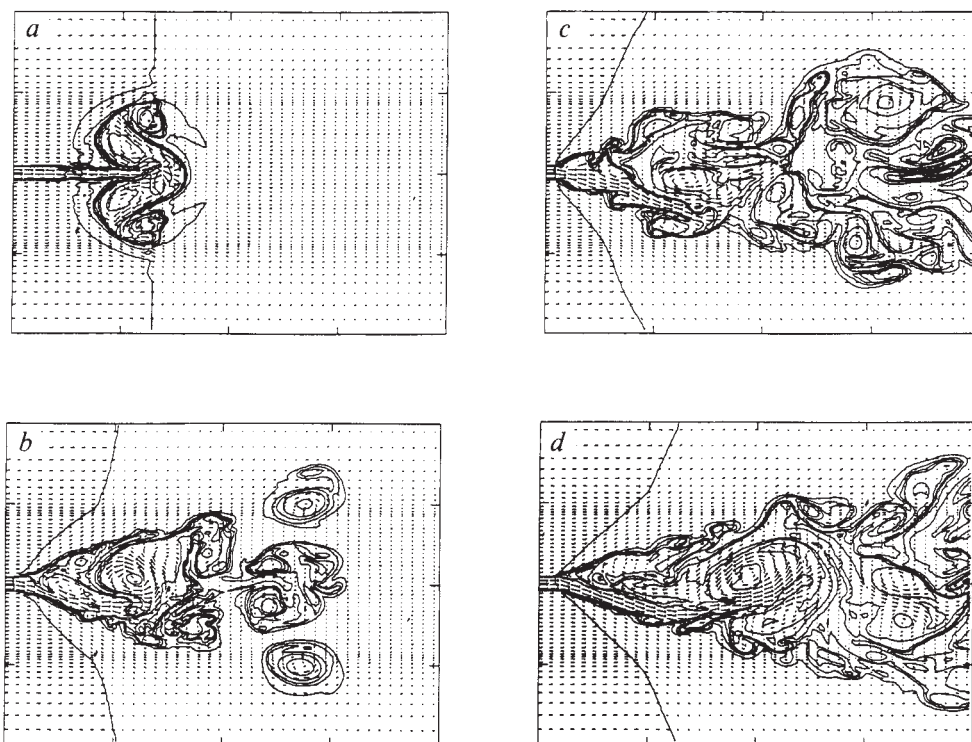


Fig. 2 Sketch of the shock-disrupted jet, showing regions and angles referred to in the text. The simulations are characterized by a series of parameters for the wind (ratio of quantity in region 2 to region 1) and for the jet (ratio of quantity in jet to region 1) as follows: pressure 10 (wind), 1 (jet); density 2.9, 0.01; temperature 3.4, 100; velocity 0.34, 8.7; sound speed 1.8, 10; Mach number 2.9, 2.5 (the last number is the ratio of jet speed to internal sound speed in region 1). The numerical simulations were performed on a Cray X-MP at the National Center for Supercomputing Applications using the ZEUS code, a descendant of the 2-d multifluid Eulerian gas dynamics code developed by Norman and Winkler²⁰ and recently extended to ideal magnetohydrodynamics²¹. Two-dimensional Cartesian geometry was used to permit growth of non-symmetric (about the mid-plane) Kelvin-Helmholtz instabilities; the jet is modelled as a propagating slab infinite in the ignorable dimension. A 1% side-to-side perturbation was applied in the injected gas stream to break lateral symmetry. Both gases are assumed adiabatic, obeying an ideal gas law with $\gamma = 5/3$. A magnetic flux loop of negligible strength is introduced with the jet gas to trace magnetic structures in the turbulent tail. The jet width W is spanned by 20 computational zones; a grid of 400×400 zones constitutes the computational domain $0 < x < 40W$, $-20W < y < 20W$. To minimize the influence of the lateral boundaries on the jet flow at reasonable computational cost, ratioed zones were used in the regions $0 < x < 40W$, $5W < |y| < 20W$; nevertheless, ~ 20 h of CPU time was needed to perform the simulation.

Fig. 3 Two-dimensional numerical simulation of a shock-disrupted initially supersonic jet at four stages of development: *a*, $t = 1 \times 10^{15}$ s; *b*, $t = 4 \times 10^{15}$ s; *c*, $t = 8 \times 10^{15}$ s; *d*, $t = 1.7 \times 10^{16}$ s (jet width = 4 kpc). Isocontours of density are superposed on velocity vectors for the slab-symmetric jet parameterized in Fig. 2. The external shock front is transmitted into the jet channel which renders the post-shock flow fully subsonic. The jet then flares to come into pressure balance with the shocked wind gas. Turbulent entrainment of ambient gas further contributes to the spreading of the jet. The large eddy seen in *d* contributes to jet deflection by creating a high stagnation pressure to one side of the jet. This deflection persists for many eddy turnover times.



A persistent post-shock deflection is recorded in the later stages of development of the flow (Fig. 3*d*). This seems to be the result of a high stagnation pressure within a large eddy which settles on one side of the jet just downstream of the external shock wave. This illustrates how easily a subsonic jet can be deflected and may help to explain how wide-angle-tailed (WAT) sources bend¹⁰. WAT jets are observed to both flare and bend at approximately the same location. Only a weak pressure gradient may be needed to bend WAT jets after they have become subsonic.

This flow disruption mechanism is insensitive to symmetry assumptions. A simulation with identical parameters in axisymmetric geometry flares and entrains ambient gas in an analogous way, although the eddy structures seem more artificial due to the imposed symmetry.

Very large array (VLA) observations of the inner lobe structure

of Cen A (NGC 5128) show (Fig. 2 of ref. 1) that the jet suddenly flares and forms a diffuse lobe in the halo of the galaxy at a distance of 4–5 kpc from the nucleus (for an assumed distance of 5 Mpc). A newly processed multiconfiguration VLA map of the northern half of the inner radio structure of Cen A is shown in Fig. 4 (D. A. Clarke and J. O. Burns, in preparation). For comparison, we also show one epoch in the numerical simulation where a large turbulent eddy has formed and has diverted the (now subsonic) jet around the eddy. The similarities in structures tempt one to interpret the lobe in Cen A as a turbulent eddy. The passive B field in the model is oriented parallel to the jet and wraps around the eddy; the projected field in Cen A follows a similar pattern in the jet and is circumferential in the lobe¹. We speculate that the last well-defined knot in the jet (at $60''$, $-20''$) could be the disrupting internal planar shock wave viewed at an oblique angle to the plane of the sky. Beyond the inner

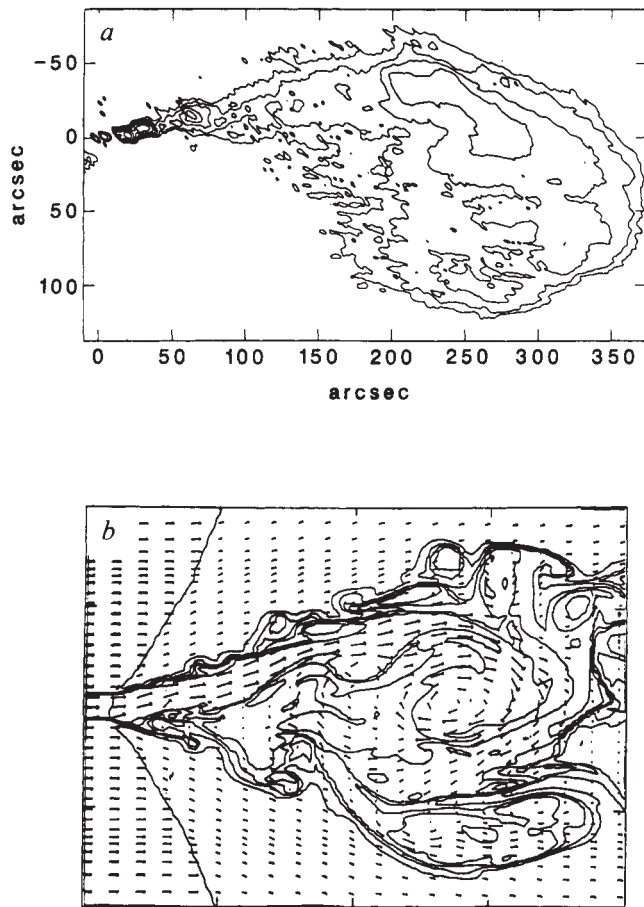


Fig. 4 A comparison of the northern inner lobe of Cen A (*a*) with a simulation (*b*), showing morphological similarity between lobe and eddy. The 5-GHz multiconfiguration VLA map by D. A. Clarke and J. O. Burns (in preparation) is rotated by 225° for ease of comparison; the compact radio core associated with the centre of the galaxy has been subtracted. Ambient density and temperature, derived from the best available radio and X-ray data, are $3 \times 10^{-4} \text{ cm}^{-3}$ and $2 \times 10^7 \text{ K}$. The wind velocity, 1060 km s^{-1} , is derived assuming an incident Mach number of 2.9 and a sound speed appropriate to gas at $2 \times 10^7 \text{ K}$. At the shock radius of 4.5 kpc the mass flow rate is taken to be $0.14 M_{\odot} \text{ yr}^{-1}$. From these numbers jet parameters are found using the ratios in Fig. 2 and assuming pressure confinement, yielding jet density 10^{-6} cm^{-3} , temperature $6 \times 10^8 \text{ K}$, velocity $9,300 \text{ km s}^{-1}$ and Mach number 2.5. These parameters are consistent with what has been inferred by other methods in other radio sources⁸.

lobes, there is a series of larger tail-like structures reaching a linear radius of 450 kpc. The most extended structure of Cen A is edge-darkened and resembles a wide-angle tail (although the galaxy environment is different than in most WAT sources).

A wind model for NGC 5128 is quite consistent with the X-ray observations¹¹. Unlike some ellipticals recently observed in Virgo, NGC 5128 has a low mass ($2 \times 10^8 M_{\odot}$), low density ($\sim 10^{-4} \text{ g cm}^{-3}$), hot ($\sim 10^7 \text{ K}$) interstellar medium. If the stars in this galaxy lose mass at a rate¹³ of $6.8 \times 10^{-13} M_{\odot} \text{ yr}^{-1}$ and the galaxy mass out to a radius of 4.5 kpc is $2 \times 10^{11} M_{\odot}$, then one should expect to observe about $1.4 \times 10^9 M_{\odot}$ built up over a timescale of 10^{10} yr , or about a factor of 10 more interstellar gas than that observed in X rays.

A partial¹² or galaxy-wide¹³ wind may be a plausible mechanism for removing gas from the interstellar medium (ISM). The stagnation radius for such a wind, found by equating the outward ram pressure of the wind with the thermal pressure of the

intergalactic medium (IGM), is

$$R_s = 4.5 \text{ kpc} \left(\frac{\dot{M}_{\text{wind}}}{0.14 M_{\odot} \text{ yr}^{-1}} \right)^{1/2} \left(\frac{n_{\text{wind},2}}{1.2 \times 10^{-4} \text{ cm}^{-3}} \right)^{1/4} \times \left(\frac{n_{\text{wind},1}}{1.2 \times 10^{-4} \text{ cm}^{-3}} \right)^{-1/4} \left(\frac{T_{\text{wind},2}}{4 \times 10^6 \text{ K}} \right)^{-1/4} \quad (2)$$

with subscripts 1 and 2 corresponding to their use in the preceding shock analysis.

The IGM pressure $n_{\text{wind},2} k T_{\text{wind},2}$, where k is Boltzmann's constant, is consistent with what is needed to confine the inner and middle lobes of Cen A and also consistent with that inferred from the presence of head-tail sources in very poor groups¹⁴. If this wind is supersonic, a shock would form near the stagnation radius where the wind meets the IGM. Such a shock appears to agree with the model initial conditions in Fig. 2, producing small changes in density and temperature (2.9 and 3.4 respectively) across the shock. This relatively weak shock would be impossible to see on the Einstein IPC map. Specific suggested parameters for the ISM/IGM and jet in Cen A that are consistent with the model and the best observations are given in the legend to Fig. 4.

The shock in the ISM of NGC 5128 could arise either from a classical wind formed by stellar and supernovae mass loss or from the active galactic nucleus¹⁵. It is interesting to note¹⁶ that the jet/lobe transition in Cen A is coincident with the innermost of a network of stellar shells¹⁷. It is unclear if such stellar shells also contain gas and thus could be interpreted as shocks¹⁵ (an alternative explanation is given by Quinn¹⁸).

The similarity in structure between WAT sources^{3,4} and the shock-disrupted jets is equally striking. Jets are observed to broaden by factors of 2–10 over a distance $< 1 \text{ kpc}$ at typical radii 10–20 kpc from the nucleus of the centrally dominant cluster galaxy (see Fig. 1). Virtually no WAT galaxies in Abell clusters have the central X-ray excess that would be taken as evidence for a cooling flow of gas. Current X-ray data are not inconsistent with a partial wind existing in these galaxies, possibly similar to those described by Loewenstein and Mathews¹².

Numerical simulations have confirmed that mildly supersonic jets can disrupt by passing through a shock in the ambient medium. If the Mach number of the jet is less than the Mach number of the external wind, a strong planar shock forms in the jet producing the transition from a well-collimated supersonic jet to a subsonic, turbulent lobe or tail. There are remarkable similarities between the basic structures of WAT and edge-darkened lobe sources with those seen in the numerical simulations. This scenario for jet disruption could potentially apply to a wide range of Fanaroff and Riley¹⁹ Class I sources. Since the shock in the external medium need not be very strong, and would thus have limited observable signatures, the radio jet/lobe transition could be a powerful probe of the ISM/IGM interface.

The numerical simulations presented here are two-dimensional and thus somewhat limited in scope. Fluid turbulence in two-dimensional cascades from smaller to larger scales unlike that seen in three-dimensional. Thus, larger scale eddies may persist longer in two-dimensional simulations than actually appear in nature. However, this is unlikely to alter the basic conclusions from the numerical simulations concerning the shock-disruption mechanism, which is our central result.

This work was supported by the US NSF. J.O.B. wishes to thank NCSA for their generous hospitality during his sabbatical visit. The authors also thank Dean Sumi for constructive discussions, and Aileen O'Donoghue for her unpublished radio map of 1231 ± 674 .

Received 23 December 1987; accepted 8 July 1988.

1. Burns, J. O., Fiegelson, E. D. & Schreier, E. J. *Astrophys. J.* **273**, 128–153 (1983).
2. Perley, R. A., Willis, A. G. & Scott, J. *Nature* **281**, 437–442 (1979).
3. Eilek, J. A., Burns, J. O., O'Dea, C. P. & Owen, F. N. *Astrophys. J.* **278**, 37–50 (1984).

4. O'Donoghue, A. & Owen, F. N. in *Radio Continuum Processes in Clusters of Galaxies* (eds O'Dea, C. D. & Uson, J. M.) 155-160 (National Radio Astronomy Observatory, Green Bank, 1986).
5. Norman, M. L., Smarr, L. L., Winkler, K.-H. A. & Smith, M. D. *Astr. Astrophys.* **113**, 285-302 (1982).
6. Norman, M. L., Winkler, K.-H. A. & Smarr, L. L. *Physics of Energy Transport in Extragalactic Radio Sources*, (eds Bridle, A. H. & Eilek, J. A.) 150-167 (National Radio Astronomy Observatory, Green Bank, 1984).
7. Shirie, J. W. & Seubold, J. G. *Am. Inst. Aeronaut. Astronaut. J.* **5**(11), 2062-2064 (1967).
8. Bridle, A. H. & Perley, R. A. *Rev. Astr. Astrophys.* **22**, 319-358 (1984).
9. Landau, L. D. & Lifshitz, E. M. *Fluid Mechanics* 336 (Pergamon, New York, 1982).
10. Burns, J. O. *Can J. Phys.* **64**, 373-378 (1986).
11. Feigelson, E. D. *et al. Astrophys. J.* **251**, 31-51 (1981).
12. Loewenstein, N. & Mathews, W. G. *Astrophys. J.* **319**, 614-631 (1987).
13. Mathews, W. G. & Baker, J. *Astrophys. J.* **170**, 241-259 (1971).
14. Stocke, J. T. & Burns, J. O. *Astrophys. J.* **319**, 671-682 (1987).
15. Williams, R. E. & Christiansen, W. A. *Astrophys. J.* **291**, 80-87 (1985).
16. Gopal-Krishna & Saripalli, L. *Astr. Astrophys.* **141**, 61-66 (1984).
17. Malin, D. F., Quinn, P. J. & Graham, J. A. *Astrophys. J.* **272**, L5-L7 (1983).
18. Quinn, P. J. *Astrophys. J.* **279**, 596-609, (1984).
19. Fanaroff, B. L. & Riley, J. M. *Mon. Not. R. astr. Soc.* **167**, 31-41 (1974).
20. Norman, M. L. & Winkler, K.-H. A. (eds) *Astrophysical Radiation Hydrodynamics*, 187-221 (Reidel, Dordrecht, 1986).
21. Clarke, D. A. thesis, Univ. New Mexico (1988).

Milliarcsecond maps of 12 GHz methanol masers

R. P. Norris*, W. H. McCutcheon*, J. L. Caswell*,
K. J. Wellington*, J. E. Reynolds†, R. S. Peng‡
& M. J. Kesteven*

*CSIRO Division of Radiophysics, P.O. Box 76, Epping,
NSW 2121, Australia

†Mount Stromlo and Siding Spring Observatories, Private Bag,
Woden, ACT 2606, Australia

‡Department of Physics, University of New South Wales,
Sydney, Australia

§Department of Physics, University of British Columbia,
Vancouver, Canada

Extremely strong methanol masers at 12 GHz have recently been discovered in sites of star formation. Here we report the first maps of these masers, which were obtained using a 275-km baseline radio-linked interferometer. The region occupied by methanol masers appears to have a similar extent to that of the well studied OH masers, and presumably surrounds the same newly formed stars. But individual clusters appear more compact than their OH counterparts, and, in contrast to the OH, appear in several cases to be distributed along a line or curve, suggesting accretion disks, shock fronts, or jets. This suggests that the methanol masers require conditions different from those that foster OH masers, and their detailed distribution promises to cast further light on the kinematics of the star-forming gas.

OH and H₂O masers are frequently found in the cool molecular gas surrounding newly formed stars, and have been used extensively to probe the physical conditions and kinematics of this gas¹. Numerous very large array, multi-element radio-linked interferometer network and very-long-baseline interferometry results have shown them to consist of clusters of small maser components scattered around the periphery of compact H II regions. In the case of OH, the individual maser components have typical sizes of 10¹² m, and are arranged in clusters typically 10-30 mpc, or 3 × 10¹⁴-10¹⁵ m in diameter.

Recently, Batrla *et al.*² discovered high-gain methanol masers (in the 2₀-3₋₁E transition at 12.178 GHz) in several sites of star formation. This species of maser displays an intensity and complexity similar to that of OH and H₂O masers, and therefore promises to be a valuable astrophysical tool for the study of star formation. Recent surveys^{3,4} have shown these methanol masers to be common in regions of star formation, and the corresponding absorption in cool clouds appears⁵ to be a powerful tracer of molecular clouds. But none of these observations has shown the detailed spatial distribution of the masers.

The first interferometric observations of these methanol masers⁶ have shown the individual maser components to have

sizes comparable to typical OH maser components. Here we present results of more extensive interferometric observations which have enabled the spatial distribution of these masers to be determined for the first time.

The observations were made during the period 28 April-1 May 1988 on the Parkes-Tidbinbilla Interferometer⁷, which has a 275-km baseline stretching from the 64-m antenna at Parkes to the 70-m NASA antenna at Tidbinbilla, giving a 20 mas fringe spacing at 12.178 GHz. Signals are transferred from Tidbinbilla to Parkes over a microwave link, and are correlated in the 1,024-channel Parkes correlator to obtain a frequency resolution of 256 complex spectral channels. The observations described here were made using a single band with a bandwidth of either 0.2, 0.5 or 1 MHz. Both antennas used linearly polarized feeds and uncooled field effect transistor receivers producing system temperatures of ~200 K, and local oscillators were derived from a hydrogen maser (at Tidbinbilla) and a rubidium standard (at Parkes). Each source was observed for typically 6 h, and in some cases observations of two nearby sources were interleaved in time. The interferometer delay and clock drift were regularly calibrated by observing unresolved continuum sources. Because the interferometer has a coherence time of only a few minutes at 12 GHz, it was not possible to derive absolute positions for the methanol masers.

The data were processed using a phase reference technique in which the phase of one strong spectral feature was subtracted from the phases of all the other spectral features, thus removing the effect of atmospheric and instrumental phase variations. The resulting phase, for an unresolved point source, is a simple sinusoidal function of hour angle. Thus the amplitude and phase of a sine curve fitted to the phase data yields parameters from which the positional offset of the feature from the reference feature may be calculated. It should be noted that these results contain no flux density information other than that given by the cross-power spectra⁶.

The phase reference technique was applied to every spectral channel containing maser emission. When two or more spatially distinct maser components emit at the same frequency, their phases interfere, and subtraction of the fitted curve from the data leaves systematic residuals. In this case the phase referencing technique becomes unreliable, and such confused channels were rejected from further analysis.

Several channels around the centre frequency of each maser feature generally gave similar positions without systematic residuals. These positions were averaged to give the quoted position for the maser component, with an estimated standard error of 3 mas in most cases. In a few cases, a better fit was obtained to one side of the peak of the spectral feature. As a check on possible systematic errors, observations of two sources were repeated on different dates using different observing configurations and with artificially induced instrumental errors. The resulting maps still agreed within the quoted errors.

Table 1 Sources observed

Source	RA (1950)	Declination (1950)	Kinematic distance (kpc)	Ref.
188.95+0.89*	06 05 54.0	21 38 40	?	4
309.92+0.48	13 47 12.7	-61 20 30	5.5	3
323.74-0.25	15 27 49.8	-56 20 15	3.0	10
331.28-0.20	16 07 40.5	-51 34 32	4.8	3
339.88-1.26	16 48 24.6	-46 03 30	3.0	3
345.01+1.79	16 53 18.0	-40 09 30	2.3	3
351.41+0.64†	17 17 32.3	-35 44 04	1.7	2

We have assumed the near kinematic distance for 331.28-0.20, but the far kinematic distance (10.0 kpc) is also consistent with the available data. In all other cases, the kinematic distance is constrained uniquely by available data.

* Sh247. † NGC6334.

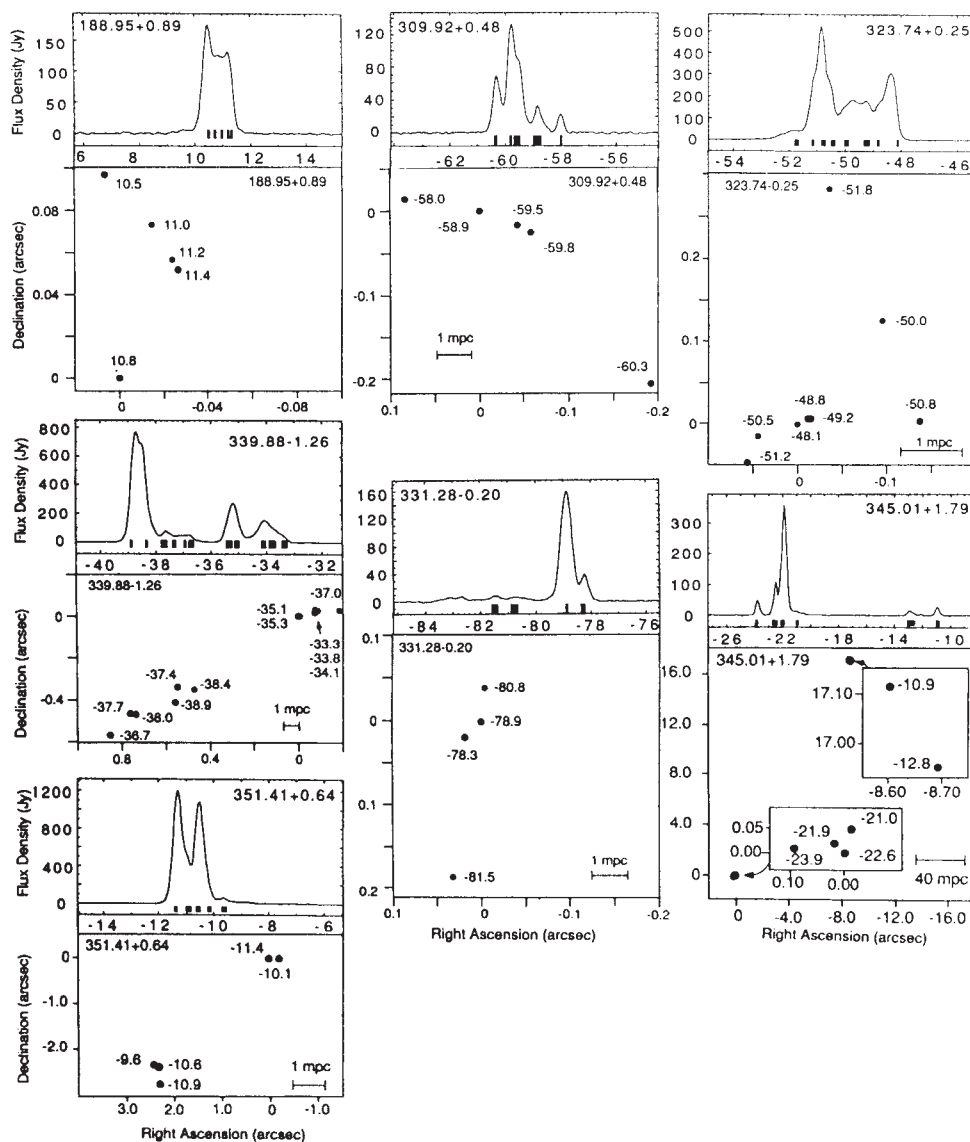


Fig. 1 Maps of the seven sources, together with single dish spectra taken on the Parkes telescope in March–May 1988. Positions are measured relative to a reference component plotted at (0, 0) in each map. We were unable to measure absolute positions with the interferometer. Standard errors in position are typically 3 mas, with the exception of 188.95+0.89, for which the standard error in declination is estimated to be 15 mas. Numbers next to the components and under the spectra are velocities (relative to the local standard of rest) in km s^{-1} . Bars on the spectra indicate the approximate position and width of the groups of channels for which positions were measured. Note that the inset boxes in the map of 345.01+1.79 are scaled by a factor of 40 with respect to the main map.

We observed a selection of seven sources, listed in Table 1, which we knew from our previous observations^{3,6} contained strong (>100 Jy) features that were substantially unresolved. The resulting maps are shown in Fig. 1. In all cases except 345.01+1.79, the maser components appear distributed over an area of less than 15 mpc (typically 0.2 arcsec) in extent, and in several cases the masers appear to be distributed in smaller clusters within these areas. 345.01+1.79 contains two such clusters separated by about 20 arcsec. The separation and orientation of the clusters in 345.01+1.79 is confirmed by single-dish measurements (J.L.C. *et al.*, manuscript in preparation) and appears to reflect the OH and H_2O maser positions⁸.

The methanol masers lie in clusters ranging in size from 1 to 10 mpc. These sizes are rather smaller than those observed in OH masers⁸, and may indicate that the methanol masers lie closer to the exciting star. But we cannot rule out the possibility that this may be a selection effect resulting from the difference in the spatial resolution between OH and methanol observations, or from the absence of the confused maser features from the methanol maps.

In several cases (notably 309.92+0.48 and 339.88-1.26), the clusters take the form of a line of maser components, whereas in 309.92+0.48 the velocities of the components change monotonically along the line. We are confident that this is not an instrumental artifact because: (1) the positions of individual channels within each feature agree within the errors; (2) the

positions of components along the lines differ by many interferometer lobes and the residuals are not systematic; and (3) 309.92+0.48 was observed twice using different configurations and at different epochs: the maps agreed within the errors.

The distribution of methanol masers along lines is unexpected, and has not previously been observed to any significant extent in other molecular masers, although we note that a similar result has been obtained⁹ for the 23.1 GHz transition of methanol. These alignments cannot readily be attributed to chance because of the large number of maser components involved, and because of the velocity gradient in 309.92+0.48. We speculate that these linear structures may be evidence for remnant accretion disks or proto-planetary systems around the newly formed stars, or collimated outflow or jets from the stars, or a shock front. Observations of additional sources will serve to determine the abundance and nature of these lines, and thus distinguish between these hypotheses.

The methanol masers in these sources cover a similar velocity range to the corresponding OH masers, but there is no detailed correspondence between individual features, and so we do not expect the detailed spatial distribution of the methanol masers to agree with that of OH. Maps of OH and H_2O masers are available⁸ for only three of the sources discussed here (339.88-1.26, 345.01+1.79 and 351.41+0.64). The lack of accurate absolute positions for the methanol masers precludes a detailed comparison, but it appears that in some cases (345.01+1.79, for

example), the relative separation of clusters of OH and H₂O masers is similar to that between methanol clusters. We deduce from this that the region occupied by the methanol masers has a similar extent to that of the OH, although differing in detail. Specifically, the methanol clusters tend to be smaller than the OH clusters and in some cases show prominent linear structures.

Although the methanol masers are distributed over an area similar to that occupied by OH masers, and cover similar velocity ranges, it appears that their detailed distribution differs significantly from that of the OH. We conclude from this that they sample a different physical regime from the OH masers. In particular, it appears from their linear distribution that in several cases they are confined to some region of special conditions, and so further observations should increase insight into the differing conditions of these species and possibly into the kinematics of star formation.

We thank the staff of the Parkes observatory and of the NASA Deep Space Tracking Station at Tidbinbilla for help during these observations.

Received 6 June; accepted 7 July 1988.

1. Elitzur, M. *Rev. mod. Phys.* **54**, 1225–1229 (1982).
2. Batrla, W., Matthews, H. E., Menten, K. M. & Walmsley, C. M. *Nature* **326**, 49–51 (1987).
3. Norris, R. P., Caswell, J. L., Gardner, F. F. & Wellington, K. J. *Astrophys. J. Lett.* **321**, L159–L162 (1987).
4. Koo, B. C., Williams, D. R. W., Heiles, C. & Backer, D. C. *Astrophys. J.* **326**, 931–940 (1988).
5. Whiteoak, J. B. *et al. Mon. Not. R. astr. Soc.* (in the press).
6. McCutcheon, W. H. *et al. Astrophys. J. Lett.* (in the press).
7. Norris, R. P., Kesteven, M. J., Wellington, K. J. & Batty, M. J. *Astrophys. J. Suppl.* **67**, 85–91 (1988).
8. Forster, J. R. & Caswell, J. L. *Astr. Astrophys.* (in the press).
9. Menten, K. M., Johnston, K. J., Wadiak, E. J., Walmsley, C. M. & Wilson, T. L. *Astrophys. J. Lett.* (in the press).
10. Kembell, A. J., Gaylard, M. J. & Nicolson, G. D. *Astrophys. J.* (in the press).

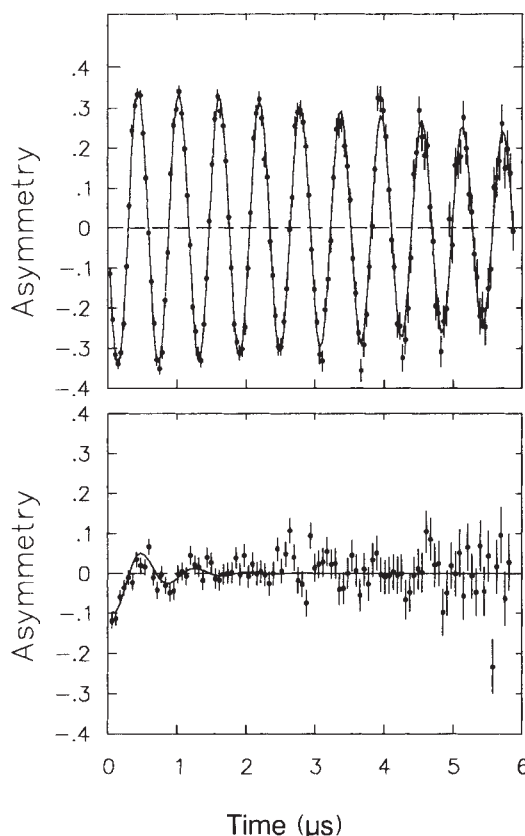


Fig. 1 Muon spin precession signal observed *a*, in a non-superconducting specimen of (Ba, K)BiO₃ at $T = 3.4$ K by applying a transverse external magnetic field $H_{\text{ext}} = 125$ G, and *b*, in an antiferromagnetic YBa₂Cu₃O_{6.18} at $T = 276$ K with $H_{\text{ext}} = 100$ G. The muon decay-time histogram is given as $N(t) = N(0) \exp(-t/\tau_\mu)[1 + AG(t) \cos(\omega t + \phi)]$, where $\tau_\mu = 2.2 \mu\text{s}$ is the muon lifetime, $A \approx 0.3$ is the initial asymmetry and $G(t)$ represents the muon spin depolarization. The asymmetry term $AG(t) \cos(\omega t + \phi)$ is shown in the figure. In μSR measurements with a transverse external field H_{ext} spectra like *a* are observed in non-magnetic or paramagnetic systems, whereas the rapid damping of the oscillating signal in *b* is characteristic of systems with static magnetic order.

Absence of magnetic order in (Ba, K)BiO₃

Y. J. Uemura*, B. J. Sternlieb*, D. E. Cox†, J. H. Brewer‡, R. Kadono‡, J. R. Kempton‡, R. F. Kiefl‡, S. R. Kreitzman‡, G. M. Luke‡, P. Mulhern‡, T. Riseman‡, D. L. Williams‡, W. J. Kossler§, X. H. Yu§, C. E. Stronach||, M. A. Subramanian¶, J. Gopalakrishnan¶ & A. W. Sleight¶

* Columbia University, New York, New York 10027, USA

† Brookhaven National Laboratory, Upton, New York 11973, USA

‡ TRIUMF and University of British Columbia, Vancouver, British Columbia V6T 2A3, Canada

§ College of William and Mary, Williamsburg, Virginia 23185, USA

|| Virginia State University, Petersburg, Virginia 23803, USA

¶ Du Pont Experimental Station, Wilmington, Delaware 19898, USA

The recent discovery^{1,2} of superconductivity in the three-dimensional perovskite system (Ba, K)BiO₃ has added a new group of compounds to the family of high-transition-temperature (high- T_c) superconductors. (Ba, K)BiO₃ contains a three-dimensional BiO network rather than the CuO planes found in two-dimensional high- T_c systems such as YBa₂Cu₃O₇, (La_{1.85}Sr_{0.15})CuO₄, Bi–Ca–Sr–Cu–O and Ti–Ba–Ca–Cu–O. There are, however, a few key features shared by (Ba, K)BiO₃ and the two-dimensional systems; for example the parent compounds BaBiO₃ and La₂CuO₄ both have one valence electron per Bi (or Cu) and doping by K (or Sr) introduces holes into the system as charge carriers. To explore the possible effects of magnetism in the (Ba, K)BiO₃ system and its related compounds, we have performed muon spin rotation (μSR) experiments on BaBiO₃, (Ba, K)BiO₃ and Ba(Bi_{0.9}Pb_{0.1})O₃, the latter being a member of another oxide superconductor system, Ba(Pb, Bi)O₃ (ref. 3), to search for static magnetic order. In the present study, no signature of magnetic order was found in any of the above systems above $T = 10$ K. This is in sharp contrast to the two-dimensional systems, where static magnetic order appears in the vicinity of superconductivity.

Muon spin rotation (μSR)⁴ is a powerful method for detecting static magnetic order. The technique has been applied extensively^{5–8} to the study of the antiferromagnets La₂CuO_{4–y}, (LaSr)CuO₄ and YBa₂Cu₃O_{6.0–6.4}. To determine the volume fraction of the magnetically ordered region of a sample, a magnetic field H_{ext} is applied to the target in a direction perpendicular to the incident muon polarization. The resulting muon decay-time histograms, produced by positron telescopes in the muon precession plane, reflect the local magnetic fields seen by the implanted muons. In the magnetically ordered state of polycrystalline samples, the random orientation of the static internal field H_{int} produced by nearby moments at a given muon site broadens the distribution of local fields, so that $H_{\text{loc}} = H_{\text{ext}} + H_{\text{int}}$. When H_{ext} is larger than ~ 100 G, this broadening of H_{loc} causes a very fast depolarization of the muon spins. In La₂CuO₄, where H_{int} is about 300–400 G, rapid depolarization has been seen. Therefore, if there were static magnetic order in BaBiO₃ and its family compounds, it should be detected easily by this method.

We performed these μSR measurements at the AGS/BNL and TRIUMF muon channels. The samples of BaBiO₃ and Ba(Pb_{0.1}Bi_{0.9})O₃ were prepared by heating intimate mixtures of appropriate quantities of BaO₂, Bi₂O₃ and PbO₂ at 800 °C for 12 h in open gold trays. The (Ba_{1–x}K_x)BiO₃ samples ($x = 0.25$)

were prepared in a similar manner but using a twofold excess of KNO_3 . Initial heating at 825°C for 3 h produced a non-superconducting cubic material with $a = 4.307 \text{ \AA}$. An oxygen anneal at 475°C for 2 h produced a superconducting specimen with $T_c = 29 \text{ K}$ and a cubic cell edge of $4.327 \text{ \AA}$.

Figure 1a shows the muon spin precession signal measured with $H_{\text{ext}} = 125 \text{ G}$ in non-superconducting $(\text{BaK})\text{BiO}_3$ at $T = 3.4 \text{ K}$. The muon spin precesses with the full initial asymmetry and a very small depolarization rate of $\sim 0.1 \mu\text{s}^{-1}$. This spectrum demonstrates that no fraction of the specimen exhibited static magnetic order at this temperature. For comparison, we show in Fig. 1b a typical μSR spectrum for a specimen with static magnetic order, in this case antiferromagnetic $\text{YBa}_2\text{Cu}_3\text{O}_{6.18}$ at $T = 276 \text{ K}$ with $H_{\text{ext}} = 100 \text{ G}$.

By similar measurements, the absence of magnetic order in BaBiO_3 and $\text{Ba}(\text{Pb}_{0.1}\text{Bi}_{0.9})\text{O}_3$ above $T = 10 \text{ K}$ has been confirmed. In addition, we have performed μSR measurements on a superconducting $(\text{BaK})\text{BiO}_3$ sample with a T_c of 29 K . From the observed spectra, we found that $\sim 20\%$ of the volume becomes superconducting below $\sim 20 \text{ K}$, in good agreement with susceptibility measurements. The remaining 80% of the sample produced a very slowly relaxing signal, implying the absence of static magnetic order in the non-superconducting portion of the specimen.

The nominal valence configuration of BaBiO_3 is Ba^{2+} and $\text{O}^{2-} \times 3$. Because the Bi atom has five valence electrons (two $6s$ and three $6p$), an average of one valence electron is left at each Bi atom, making BaBiO_3 a system with a half-filled band. This material has a monoclinic crystal distortion⁹ related to the disproportionation $2\text{Bi}^{4+} \rightarrow \text{Bi}^{3+} + \text{Bi}^{5+}$. When a small amount of Pb or K is introduced into the system, this distortion is suppressed. We have made neutron and X-ray scattering measurements on the same specimens used in the present μSR study, which confirm that BaBiO_3 has a monoclinic, $\text{Ba}(\text{Pb}_{0.1}\text{Bi}_{0.9})\text{O}_3$ an orthorhombic, and $(\text{BaK})\text{BiO}_3$ a cubic crystal structure (to within the limit of instrumental resolution) below room temperature. The absence of static magnetic order in these systems is thus independent of crystal structure.

The present results do not preclude the existence of a rapidly fluctuating moment on the Bi atom. For $H_{\text{int}} \approx 300 \text{ G}$, dynamic effects can be detected by μSR only when the fluctuation time τ_c is slower than 10^{-12} s . μSR is extremely sensitive to a static local field, so that even a nuclear dipolar field can be easily measured. The present experiment has shown that the ordered moment in these systems, if one exists at all, is smaller than $\sim 0.01 \mu_B$ per formula unit.

A remote possibility still exists that static magnetic order could have escaped detection. This could be the case if (1) the ordering temperature was below our range of measurement; (2) the composition of the specimen was outside the region with magnetic order; (3) the static internal field at the muon site cancels as a result of crystal symmetry. The latter is very unlikely, especially in randomly substituted systems. We intend to expand the temperature and composition ranges to investigate possibilities (1) and (2).

The absence of magnetic order in the parent compound BaBiO_3 and systems related to the superconducting phase of $(\text{BaK})\text{BiO}_3$ is in clear contrast to the case of the two-dimensional systems $(\text{La, Sr})\text{CuO}_4$ and $\text{YBa}_2\text{Cu}_3\text{O}_x$, where magnetic order is always seen in the vicinity of superconductivity. This may be due to the fact that the $6s$ electrons of Bi are less localized than the $3d$ electrons of Cu in the two-dimensional systems. The present result demonstrates that magnetic order in the parent compound is not a necessary condition for superconductivity with $T_c \approx 30 \text{ K}$.

4. For general aspects of muon spin relaxation, see proceedings of four previous international conferences: *Hyperfine Interactions* 6 (1979); 8 (1981); 17-19 (1984); 31 (1986).
5. Uemura, Y. J. *et al.* *Phys. Rev. Lett.* 59, 1045-1048 (1987); *Physica C* 153-155, 760-770 (1988).
6. Brewer, J. H. *et al.* *Phys. Rev. Lett.* 60, 1073-1077 (1988); Nishida, N. *et al.* *J. Phys. Soc. Japan* 57, 599 (1988).
7. Budnick, J. I. *et al.* *Europhys. Lett.* 5, 651 (1988).
8. Harshman, D. R. *et al.* *Phys. Rev. B* 38, 852-855 (1988).
9. Cox, D. E. & Sleight, A. W. *Solid St. Commun.* 19, 469 (1976).

Terrestrial feedback in atmospheric oxygen regulation by fire and phosphorus

L. R. Kump

Earth System Science Center, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

The regulation of atmospheric oxygen levels (p_{O_2}) occurs on million-year timescales and is effected by modulation of sedimentary organic carbon burial and weathering rates^{1,2}. Until recently it was believed that these processes were dominated by material produced by marine organisms; now it appears that terrestrial plants contribute a significant amount of organic detritus to marine sediments. Here I explore the possibility of a coupling between the rates of terrigenous and marine sedimentary organic carbon burial (B_{OC}) that might stabilize p_{O_2} at or near the present atmospheric level. The coupling involves fire as a means to transfer phosphorus, the nutrient that limits global B_{OC} ^{3,4} and thus global oxygen production rates, from terrestrial to marine environments.

The generally accepted mechanism of oxygen control is based on the assumption that marine B_{OC} dominates the global rate^{1,2,5-7}. At levels near today's, an increase in p_{O_2} causes a decrease in marine B_{OC} (net photosynthesis = net oxygen production) because more of the global ocean floor becomes bathed in oxygenated waters. Conversely with a p_{O_2} decrease, the extent of oceanic anoxia increases, enhancing the preservation (burial) of organic matter. This general mechanism has at least three limitations: (1) it ignores the dynamics of the oceanic phosphorus (P) cycle, (2) it ignores the contribution of terrigenous organic carbon to global B_{OC} , and (3) because it is not coupled to terrestrial processes, it may not provide a sensitive feedback capable of preventing global wildfires.

Of these limitations, the marine P cycle has received the most attention in recent years⁸. The removal of P from the surface oceans with organic matter⁴ and iron hydroxides⁹ is essentially quantitative, and surface P concentrations are severely depleted. At depth and in surface sediments¹⁰, most of the organic matter settling from the surface ocean is regenerated (aerobically decomposed), leading to higher P concentrations in the deep ocean. The upwelling of deep waters then returns P to the surface, where it is once again subject to organic and inorganic scavenging. Little is known about the internal inorganic P cycle, but presumably the scavenging of P on iron hydroxides would not be reversed in settling through oxic waters with enhanced P concentrations. Thus the extent of deep-ocean regeneration, which is related to p_{O_2} , may determine the relative proportions of organic and inorganic P removal, and thus the rates of marine B_{OC} and oxygen production.

In the earlier considerations of marine oxygen control, only organic P removal was assessed. But without competition from inorganic processes, organic P removal on geological timescales would always equal the river input, and there would be no O_2 feedback. For example, an increase in p_{O_2} may cause an increase in regeneration rates, but this alone would not diminish B_{OC} because deep-ocean P concentrations, and thus P upwelling fluxes to the surface, would also increase⁴ on a timescale much shorter than the p_{O_2} response. B_{OC} would simply become a smaller fraction of surface primary productivity.

Received 1 August; accepted 8 August 1988.

1. Mattheiss, L. F., Gyorgy, E. M. & Johnson, D. W. Jr. *Phys. Rev.* B37, 3745-3746 (1987).
2. Cava, R. J. *et al.* *Nature* 332, 814-816 (1988).
3. Sleight, A. W., Gillson, J. L. & Bierstedt, P. E. *Solid St. Commun.* 17, 27-28 (1975).

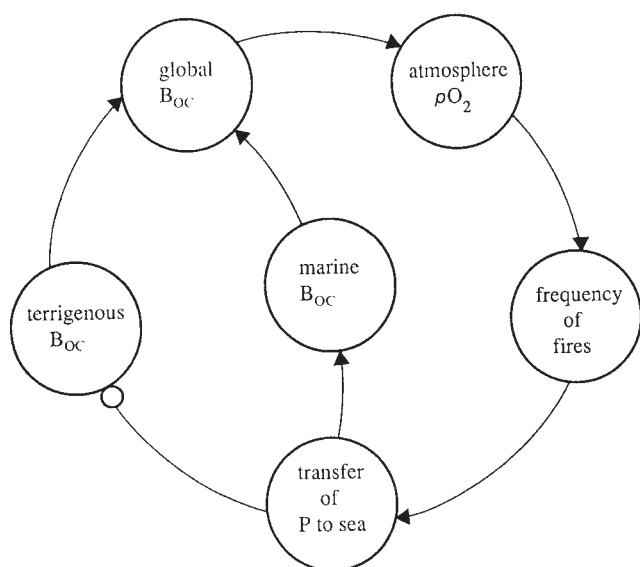


Fig. 1 Feedback loop in contemporary atmospheric oxygen regulation. Positive feedbacks are indicated by arrows, negative feedbacks by small circles. B_{OC} is the burial rate of organic C.

It has become clear that any mechanism proposed for the marine control of atmospheric p_{O_2} must also accommodate the flux of terrigenous organic carbon to sediments. Terrestrially derived organic C is now recognized as an important component of the total organic C in marine sediments (where most organic C burial takes place today)¹¹, both in nearshore and open oceanic environments. Biochemical tracers indicate that vascular plant material contributes to both particulate¹² and dissolved¹³ organic matter in the sea. The relative proportion of terrigenous to marine organic carbon burial presumably decreases offshore, and is low in high-productivity/upwelling regions where the rain of planktonic detritus is high. But Berner¹¹ has shown that nearshore environments dominate global B_{OC} , and in these environments terrigenous organic carbon is often dominant. Ittekkot's¹⁴ estimate of the riverine, refractory terrigenous organic carbon flux, much of which is likely to be buried, is of the order of 12×10^{12} mol yr⁻¹. This estimate is close to Berner's¹¹ estimate of the entire global B_{OC} . Also, for particular times in the geological past, terrigenous organic carbon has been shown to be an important component of both marine and terrestrial sediment¹⁵⁻¹⁸.

Terrigenous organic carbon poses a problem for the marine O_2 control hypothesis because of its resistance to aerobic degradation¹⁴. Increased p_{O_2} levels probably would have little effect on terrigenous B_{OC} in marine sediments, and without feedback there could be no effective regulation of atmospheric oxygen. An implication is that the evolution of land plants in the early Palaeozoic¹⁹ was probably associated with an increase in p_{O_2} , not because B_{OC} increased, and not because higher O_2 concentrations were needed to oxidize the new refractory organic matter flux, but because higher p_{O_2} was needed to prevent the burial of an amount of labile marine organic carbon equivalent to the new, terrigenous source.

Another problem with an O_2 regulation mechanism that is entirely marine is that it would not obviously prevent a slight rise of p_{O_2} from today's level. Such a rise would be sufficient to place the Earth system in continual danger of global conflagration, a condition that is inconsistent with the fossil record of terrestrial environments. Calculations by Watson *et al.*²⁰ on variable water contents demonstrated that at today's p_{O_2} , fuel near the water saturation point ($\sim 40\%$) has a very low probability of ignition. But an increase in p_{O_2} to 30% (a change that could occur in a few million years) causes a tremendous increase in the relative probability of ignition. Furthermore, for a typical

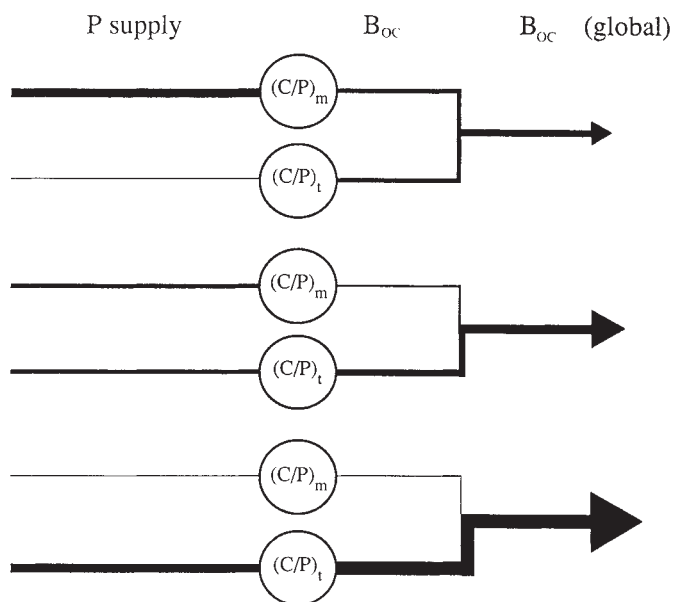


Fig. 2 The effect of the distribution of the limiting P supply between marine and terrigenous sources of organic C on global organic carbon burial rates (B_{OC}). Global B_{OC} is smallest when P is largely sequestered into marine organic C. As the proportion of terrigenous sequestering increases, so too does the global B_{OC} , because the C/P of terrestrial organic C (C/P_t) is an order of magnitude larger than that of marine organic C (C/P_m).

natural fuel with a moisture content of $\sim 10\%$, the probability of ignition rises rapidly with an increase in p_{O_2} , even at today's p_{O_2} . Thus the response of fire to today's p_{O_2} is sensitive at normal moisture contents, and there is a high likelihood of ignition with an increase of p_{O_2} , even during a downpour.

Modelling efforts²¹⁻²³ have clearly established the need for a strong feedback in oxygen regulation. Because of the problems with a marine oxygen feedback, a new hypothesis is proposed: there is a geophysiological²⁴ coupling between terrigenous and marine-produced organic carbon burial rates that is sensitive to fire and has the capacity to regulate global B_{OC} and thus oxygen-production rates. The coupling involves the distribution of phosphorus between oceanic and land ecosystems (Fig. 1). For example, an increase in p_{O_2} causes an increase in the frequency of fires. The ecological disturbances from fire cause a transfer of phosphorus (P) from land to sea. Marine B_{OC} increases, but terrigenous and global B_{OC} and O_2 production rates decrease. Thus the overall feedback is negative; a perturbation in p_{O_2} is followed by a restorative sequence of events coupled to the global cycle of P.

Fires cause a number of changes in terrestrial systems that could result in the removal of P from land biota and soils. These include enhanced soil erosion as the result of the loss of ground cover^{25,26}, the atmospheric transport of ash (generated from the P-rich leaves and stems of plants)²⁷⁻²⁹, and the prevention of the establishment of 'climax' communities that would be nutrient-efficient and prevent the loss of P from the land system^{30,31}. Even with a substantial increase in the number of fires, the loss of P would be small, because much of the transported soil and ash would be redeposited on land. It is reasonable, however, to assume that an increase in the P flux to the ocean would result.

The proposed mechanism for oxygen control hinges on the fact that terrestrial plants use P more efficiently. This higher efficiency is reflected in the larger molar C/P ratio of terrestrial vascular ($C/P \approx 2,000$) versus marine ($C/P \approx 100$) plants³². The C/P of marine organic matter may be increased during diagenesis, although there is no clear consensus on the ubiquity

of this process³³⁻³⁵. Here the assumption is that this effect is secondary. If the supply of P limits global B_{OC} , then a higher rate can be sustained when terrigenous material dominates. Mathematically, $B_{OC}(\text{global}) = W_p \cdot [f \cdot (C/P)_t + (1-f-i) \cdot (C/P)_m]$, where W_p is the global weathering rate of P, f is the fraction of W_p buried as terrigenous organic carbon and i is the fraction of W_p buried as inorganic P, $(C/P)_t$ represents C/P in terrigenous organic carbon and $(C/P)_m$ that in marine organic carbon. When more of this limiting supply of P is buried as marine organic matter, global carbon burial rates fall (Fig. 2).

The existence of the proposed feedback is not conclusively proven by the arguments and observations presented here. Nevertheless, the mechanism is reasonable in light of our current understanding of the relationship between p_{O_2} and the combustibility of wood, and of the global cycles of C and P. It is unique in that it places the dominant regulatory feedback in the terrestrial realm, where the biota are most sensitive to ecological changes wrought by fluctuations in p_{O_2} . It also suggests that pre-land plant p_{O_2} values were lower than today's (compare refs 36-37), and could have been insufficient to sustain a fire (flames require $p_{O_2} > 13\%$)³⁸. They may also have been insufficient to prevent widespread oceanic anoxia, and under these conditions the inorganic removal rate of P with iron hydroxides would have been substantially depressed. Fires and P-rich iron hydroxides may owe their existence to the evolution of terrestrial vascular plants.

I am grateful to Dr J. Kasting for a constructive review. Support for this work was provided in part by grants from the American Chemical Society, Petroleum Research Fund, and the National Science Foundation.

Received 16 June; accepted 29 July 1988.

- Walker, J. C. G. *Evolution of the Atmosphere* (Macmillan, New York, 1977).
- Holland, H. D. *The Chemistry of the Atmosphere and Oceans* (Wiley, New York, 1978).
- Redfield, A. C. *Am. Sci.* **46**, 205-222 (1958).
- Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Eldigio, Palisades, New York, 1982).
- Broecker, W. S. *J. geophys. Res.* **75**, 3553-3557 (1970).
- Walker, J. C. G. *Bioscience* **34**, 486-491 (1984).
- Holland, H. D. *The Chemical Evolution of the Atmosphere and Ocean* (Princeton University Press, New Jersey, 1984).
- Froelich, P. N., Bender, M. L., Luedtke, N. A., Heath, G. R. & DeVries, T. *Am. J. Sci.* **282**, 474-511 (1982).
- Sherwood, B. A., Sager, S. L. & Holland, H. D. *Geochim. cosmochim. Acta* **51**, 1861-1866 (1987).
- Jahnke, R. A. & Jackson, G. A. *Nature* **329**, 621-623 (1987).
- Berner, R. A. *Am. J. Sci.* **282**, 451-473 (1982).
- Gagosian, R. B., Peltzer, E. T. & Merrill, J. T. *Nature* **325**, 800-803 (1987).
- Meyers-Schulte, K. J. & Hedges, J. I. *Nature* **321**, 61-63 (1986).
- Ittekkot, V. *Nature* **332**, 436-438 (1988).
- Arthur, M. A., Dean, W. E. & Schlanger, S. O. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 504-530 (Am. Geophys. Un., Washington D.C., 1985).
- Stein, R., Rulkotter, J. & Welte, D. H. *Chem. Geol.* **56**, 1-32 (1986).
- Simoneit, B. R. T. *Mar. Geol.* **70**, 9-41 (1986).
- Berner, R. A. & Raiswell, R. *Geochim. cosmochim. Acta* **47**, 855-862 (1983).
- Beerbower, R. in *Geological Factors and the Evolution of Plants* (ed. Tiffney, B. H.) 47-91 (Yale University Press, 1985).
- Watson, A., Lovelock, J. E. & Margulis, L. *Biosystems* **10**, 293-298 (1978).
- Lasaga, A. C., Berner, R. A. & Garrels, R. M. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 397-411 (Am. Geophys. Un., Washington D.C., 1985).
- Kump, L. R. & Garrels, R. M. *Am. J. Sci.* **286**, 337-360 (1986).
- Berner, R. A. *Am. J. Sci.* **287**, 177-196 (1987).
- Lovelock, J. E. in *Geophysiology of Amazonia, Vegetation and Climate Interactions* (ed. Dickinson, E.) 526 pp. (Wiley, New York, 1987).
- Rowe, P. B. *Trans. Am. geophys. Un.* **25**, 84-86 (1944).
- Hingston, F. J. & Raison, R. J. in *Cycling of C, N, S and P in Terrestrial and Aquatic Ecosystems* (eds Freney, J. R. & Galbally, I. E.) 11-24 (Springer, Berlin, 1982).
- Lloyd, P. S. *J. Ecol.* **59**, 261-273 (1971).
- Smith, D. W. & Bowes, G. C. *Can. J. Soil Sci.* **54**, 214-224 (1974).
- Harwood, C. E. & Jackson, W. D. *Aust. Forest.* **38**, 92-99 (1975).
- Krebs, C. J. *Ecology: The Experimental Analysis of Distribution and Abundance* 2nd edn, 511 (Harper & Row, New York, 1978).
- Knoll, M. A. & James, W. C. *Geology* **15**, 1099-1102 (1987).
- Likens, G. E., Bormann, F. H. & Johnson, N. M. in *Some Perspectives of the Major Biogeochemical Cycles* (ed. Likens, G. E.) 93-122 (Wiley, New York, 1981).
- Hartmann, M., Muller, P. J., Suess, E. & van der Weijden, C. H. *Meteor.-Forschungsergebnisse* **C24**, 1-67 (1976).
- Nissenbaum, A. *Geochim. cosmochim. Acta* **43**, 1973-1978 (1979).
- Mach, D. L., Ramirez, A. & Holland, H. D. *Am. J. Sci.* **287**, 429-441 (1987).
- Rhoads, D. C. & Morse, J. W. *Lethaia* **4**, 413-428 (1978).
- Cope, M. J. & Chaloner, W. G. in *Geological Factors and the Evolution of Plants* (ed. Tiffney, B. H.) 257-277 (Yale University Press, 1985).
- Drysdale, D. *An Introduction to Fire Dynamics* (Wiley, New York, 1985).

Remote sensing of canopy chemistry and nitrogen cycling in temperate forest ecosystems

Carol A. Wessman^{*||}, John D. Aber^{†¶},
David L. Peterson[‡] & Jerry M. Melillo[§]

^{*} Department of Forestry and the Environmental Remote Sensing Center, University of Wisconsin, Madison, Wisconsin 53706, USA

[†] Department of Forestry, University of Wisconsin, Madison, Wisconsin 53706, USA

[‡] Ames Research Center, National Aeronautics and Space Administration, Moffett Field, California 94035, USA

[§] Ecosystems Center, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

Foliar lignin content (referring to acid-insoluble fibre in foliage) is a primary rate-limiting factor of forest-litter decomposition and is thus important in modulating nutrient-cycling rates in forest ecosystems. Here we report the use of images acquired by the Airborne Imaging Spectrometer, an experimental high-spectral-resolution imaging sensor developed by NASA, to estimate the lignin concentration of whole forest canopies in Wisconsin. The observed strong relationship between canopy lignin concentration and nitrogen availability (through nitrogen mineralization) in seven undisturbed forest ecosystems on Blackhawk Island, Wisconsin, suggests that canopy lignin may serve as an index for site nitrogen status. This predictive relationship presents the opportunity to estimate nitrogen-cycling rates across forested landscapes through remote sensing.

Remote sensing is used increasingly for the measurements required to develop landscape, regional and global assessments of the state of the biosphere. To date, most applications of remote sensing to terrestrial ecosystems have involved the estimation of foliar area and biomass, or absorbed photosynthetically active radiation¹⁻³. None has been successful in predicting rates of ecosystem processes occurring in soils. Here we present the first accurate estimation of lignin content of whole forest canopies by remote sensing. In one set of undisturbed temperate forest ecosystems, a strong correlation between lignin content and nitrogen mineralization allows mapping of nitrogen cycling rates.

Research sites included 18 well-studied forest stands; disturbed or revegetated sites were located in the University of Wisconsin Arboretum in Madison, and undisturbed sites on Blackhawk Island, Wisconsin, a large island in the Wisconsin River ~70 km north of Madison. Thirteen of the stands are composed of mixed hardwoods dominated by species of oak and maple. Pure and mixed stands of red and white pine make up the remaining five. Previous measurements on these sites include net primary productivity, nitrogen mineralization (determined by *in situ* soil incubations) and full soil characterizations^{4,5}.

Total canopy biomass and chemistry values were determined by (1) collection of leaf litterfall in all stands for estimation of foliar biomass (values for pine stands were increased by a measured foliage retention rate on twigs); (2) late-summer pre-senescence collection of replicate, composite random samples of green foliage in each stand; (3) measurement of total lignin (sulphuric acid digestion) and nitrogen (micro-Kjeldahl) concentration of foliage samples; (4) multiplication of foliar concentration and biomass values for canopy-chemistry totals⁶.

Of the three canopy characteristics measured, total nitrogen

^{||} Present address: Center for the Study of Earth from Space, Cooperative Institute for Research in Environmental Sciences, University of Colorado/NOAA, Boulder, Colorado 80309-0449, USA.

[¶] Present address: Complex Systems Research Center, EOS, University of New Hampshire, Durham, New Hampshire 03824, USA.

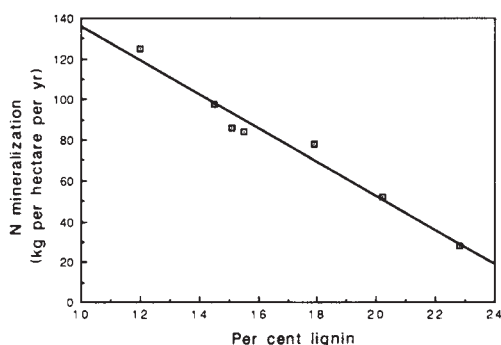


Fig. 1 Per cent canopy lignin versus annual nitrogen mineralization ($N_{\min}$) for 7 undisturbed sites on Blackhawk Island, Wisconsin. $R^2 = 0.96$, $n = 7$; $N_{\min} = 219.4 - 8.3 \times \% \text{ lignin}$.

showed the smallest range of values (64–111 kg per hectare, coefficient of variation (cv) = 0.15), despite the wide variation in nitrogen availability to plants through mineralization (28–135 kg nitrogen per hectare per year, cv = 0.40). Canopy biomass showed a wider range (2.92–8.46 tons per hectare, cv = 0.37), mainly as a result of high biomass values in the pine stands. Total lignin content was still more variable (0.197–1.79 tons per hectare, cv = 0.66), because lignin concentrations vary widely between deciduous species and are highest in pine foliage. Lignin concentration is very highly correlated with nitrogen cycling rates for the subset of undisturbed sites (Fig. 1), reflecting the importance of lignin in determining rates of litter decomposition and the long-term formation and decomposition of soil organic matter⁷.

Remote-sensing data were acquired using NASA's Airborne Imaging Spectrometer (AIS), a prototype instrument using area detector arrays to generate high-spectral-resolution images⁸. The AIS images 128 contiguous spectral bands of reflected radiation between 0.9 and 2.4 μm at 9.3 nm spectral and $11 \times 11 \text{ m}$ spatial (at 6 km altitude) resolution. Preprocessing of the Wisconsin AIS data included correction of inter-detector response variations and the application of a notch filter to the Fourier transform of the imagery to remove vertical and horizontal striping⁹. Spectral calibration of the data was approximated using a post-flight calibration record and by comparison of broad spectral features with a LOWTRAN 6 simulation of atmospheric transmission for a standard mid-latitude summer¹⁰. Instrument problems limited our analysis to 1.2–1.6 μm (ref. 11).

The spectra from a 3×3 pixel area covering each research site were averaged to reduce effects of local canopy heterogeneity and were compared with canopy data (field-data plots were $30 \times 30 \text{ m}$ in size). A first-order differentiation of the spectra, defined as the difference in signal between adjacent spectral bands and referred to as 'first difference', was performed to reduce broad brightness variations while emphasizing changes of spectral slope that could be related to absorption features¹². Hence first-order baseline shifts, which could be due to canopy architecture, atmospheric differences, solar insolation and sensor effects, are also reduced.

Stepwise multiple regression was used to identify those combinations of wavelengths that were most highly correlated with canopy chemistry and biomass. There was strong correlation

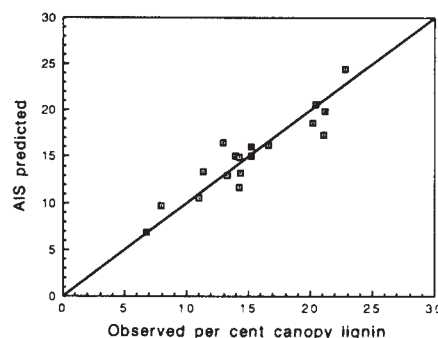


Fig. 2 Per cent canopy lignin observed in the field for 18 forest stands versus AIS predicted values using three first difference bands: 1,256, 1,555, 1,311 nm. $R^2 = 0.85$, $n = 18$.

between lignin concentration (content per biomass) and selected bands of the transformed AIS data across all deciduous and coniferous stands. As subset regression can lead to inflated correlation coefficients (R) when the number of predictor variables exceeds the numbers of samples, approximation formulae derived by Rencher and Pun¹³ were used as guidelines for assessing the significance of the R^2 values. The estimation equation used here for canopy lignin concentration ($R^2 = 0.85$, standard error = 1.9, significant at 0.05 level) is based on three terms (first difference values at 1,256, 1,555 and 1,311 nm) (Table 1, Fig. 2). Correlations with total canopy biomass or nitrogen content were less significant¹⁴.

Further research is required to understand the physical significance of the empirically selected bands. We suggest that absorption at 1,256 nm corresponds to C–H bond vibrations in organic constituents which include or co-vary with lignin. Absorption in the 1,500–1,600 nm region has been related to water content of plant canopies¹⁵ and to lignin content of dried ground leaf samples^{16,17}. Absorption at 1,311 nm, the limb of the 1,450 nm absorption peak from water and O–H bonds, is rapidly varying, and, in the transformed data, is probably related to canopy water and O–H bonds in lignin¹⁸. The actual influence, and thus detection sensitivity by remote sensing, of lignin on absorption at each of these wavelengths, particularly the latter two, depends on the interrelationship between lignin and water absorption properties, as well as the signal-to-noise ratio in the spectral data.

The lignin spectrum has not been characterized adequately because of difficulties in extracting lignin (an amorphous polymer) from foliar tissue in a pure, isolated form¹⁹. Consequently, it is not possible to calculate the influence of lignin on vegetation spectra in the presence of other scattering and absorbing constituents (particularly water). Additional experimental data collection and theoretical investigations are required to confirm the interpretation of the imaging spectrometer results presented here. Research has been initiated recently to address questions regarding the covariance of leaf constituents and their specific absorptivities. Radiative transfer modelling will be instrumental in this respect.

Field measurements, made in 1984, of nitrogen mineralization and canopy lignin content for three sites on Blackhawk Island are available to test the relationships described above. Nitrogen

Table 1 Input variables and statistics for multiple regression analysis of per cent canopy lignin

Step	R^2	Standard error of estimate	y-intercept	Coefficients of variables		
				1,256 nm	1,555 nm	1,311 nm
0		4.6	15.13			
1	0.71	2.6	33.38	0.18		
2	0.82	2.1	34.33	0.14	−0.25	
3	0.85	1.9	34.07	0.20	−0.27	−0.03

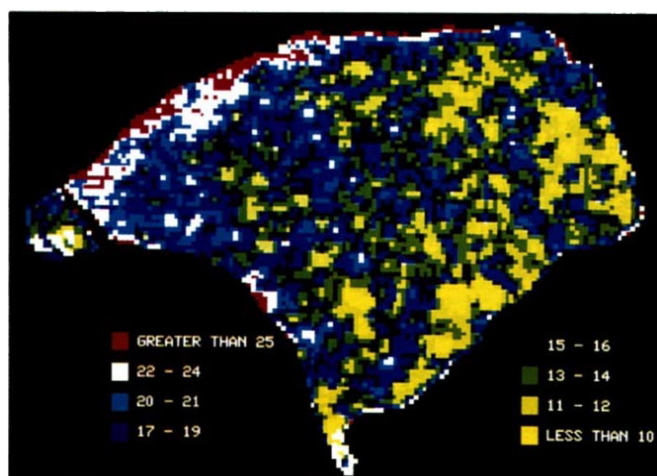


Fig. 3 Spatial distribution of canopy lignin concentration across Blackhawk Island. The image is a mosaic of six flightlines. Changes in per cent lignin from left to right across the island reflect a continuous change in soil texture resulting from sediment sorting when the island area was an early post-glacial floodplain.

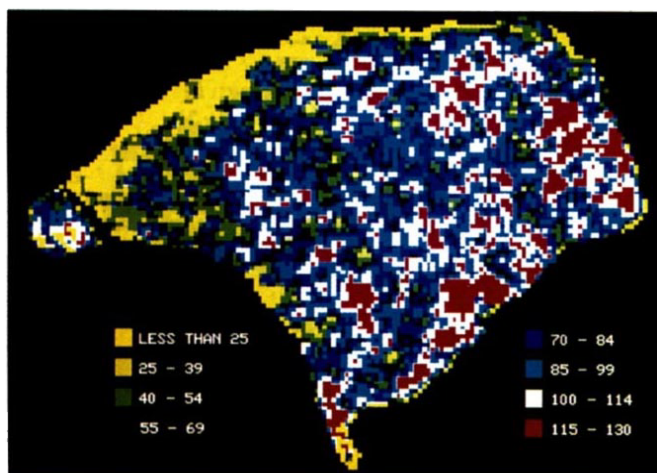


Fig. 4 Annual nitrogen mineralization for Blackhawk Island, Wisconsin, as estimated from AIS-predicted canopy lignin concentrations.

mineralization has also been measured on a fourth site. Using AIS spectra for those sites, predicted per cent lignin for the three stands were 18.2%, 17.0% and 14.6% versus the respective values of 15.5%, 15.1% and 14.4% measured in 1984²⁰. These values approximate the observed precision ($\leq 10\%$) with which canopy lignin content can be measured and the expected year-to-year variation in lignin content. Nitrogen-mineralization predictions were 67, 77, 98 and 25 kg nitrogen per hectare per year versus the measured 1984 values of 84, 86, 98 and 29 kg nitrogen per hectare per year, respectively.

AIS data were used in the regression equations derived above to produce a map of canopy lignin concentration for Blackhawk Island (Fig. 3), and thus, using the relationship in Fig. 1, a map of annual rates of nitrogen mineralization (Fig. 4). The first image depicts a continuous spatial distribution of per cent lignin values that correlates with the distribution of species with high- and low-lignin foliage⁴. Areas with per cent lignin values higher than 19% are dominated by pines, those below 13% are dominated by sugar maple and basewood, and those in between are dominated by oaks.

The second image is, as far as we know, the first published successful attempt to predict an important biogeochemical pro-

cess in forest ecosystems through remote sensing. Although the relationship depends on the close correlation between canopy composition and nitrogen mineralization, and therefore is not generalizable to seriously disturbed sites, it could have broad application to managed and natural forests and could allow prediction of nitrogen cycling rates at the landscape or regional level.

We thank S. McNulty and S. Kaat for field work, D. Card, K. Wolter and B. Yandell for statistical advice, L. Strong for assistance with image processing, and Drs C. Tucker and O. Faix for helpful comments on the manuscript. This research was supported by NASA under a Joint Research Interchange with Ames Research Center, NSF and the McIntire/Stennis Program through the College of Agriculture and Life Sciences, University of Wisconsin-Madison.

Received 16 September 1987; accepted 28 July 1988.

1. Tucker, C. J., Fung, I. Y., Keeling, C. D. & Gammon, R. H. *Nature* **319**, 195-197 (1986).
2. Peterson, D. L., Spanner, M. A., Running, S. W. & Teuber, K. B. *Remote Sens. Environ.* **22**, 323-341 (1987).
3. Asrar, G., Fuchs, M., Kanemasu, E. T. & Hatfield, J. L. *Agron. J.* **76**, 300-304 (1984).
4. Pastor, J., Aber, J. D., McClaugherty, C. A. & Melillo, J. M. *Ecology* **65**, 256-268 (1984).
5. Nadelhoffer, K. J., Aber, J. D. & Melillo, J. M. *Plant and Soil* **80**, 321-324 (1984).
6. McClaugherty, C. A., Pastor, J., Aber, J. D. & Melillo, J. M. *Ecology* **66**, 266-275 (1985).
7. Melillo, J. M., Aber, J. D. & Muratore, J. F. *Ecology* **63**, 621-626 (1982).
8. Goetz, A. F. H., Vane, G., Solomon, J. E. & Rock, B. N. *Science* **228**, 1147-1153 (1985).
9. Hlavka, C. *Proc. Second AIS Data Analysis Workshop* (JPL Publ. 86-85) 74-78 (1986).
10. Wessman, C. A., Aber, J. D. & Peterson, D. L. *Proc. Int. Geosci. Remote Sens. Symp.* **2**, 1189-1192 (1987).
11. Conel, J. E., Adams, S., Alley, R. E., Hoover, G. & Shultz, S. *Proc. Second AIS Data Analysis Workshop* (JPL Publ. 86-85) 31-39 (1986).
12. Dixit, L. & Ram, S. *Appl. Spectrosc. Rev.* **21**, 311-418 (1985).
13. Rencher, A. C. & Pun, F. C. *Technometrics* **22**, 49-53 (1980).
14. Wessman, C. A. thesis, Univ. Wisconsin, Madison (1987).
15. Ripple, W. J. *Photogr. Engng Remote Sens.* **52**, 1669-1675 (1986).
16. Norris, K. H., Barnes, R. F., Moore, J. E. & Shenk, J. S. *J. Anim. Sci.* **43**, 889-897 (1976).
17. Card, D. H., Peterson, D. L., Matson, P. A. & Aber, J. D. *Remote Sens. Environ.* (in the press).
18. Ziehm, W. *Geochim. cosmochim. Acta* **28**, 1555 (1964).
19. Hergert, H. L. in *Lignins: Occurrence, Formation, Structure and Reactions* (eds Sarkanen, K. V. & Ludwig, C. H.) 267-297 (Wiley-Interscience, New York, 1971).
20. Fownes, J. H. thesis, Univ. Wisconsin, Madison (1985).

A silica-rich sodium pyroxene phase with six-coordinated silicon

R. J. Angel*, T. Gasparik†, N. L. Ross*, L. W. Finger*, C. T. Prewitt* & R. M. Hazen*

*Geophysical Laboratory, Carnegie Institution of Washington, 2801 Upton Street NW, Washington, DC 20008, USA

†Department of Earth and Space Sciences, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

Phase transitions in minerals and melts that involve a change in the coordination of silicon from fourfold to sixfold are generally accompanied by large changes in material properties such as density, bulk modulus and elastic moduli. When such transformations occur within the Earth, these large changes in material properties can result in seismic discontinuities. The presence of six-coordinated silicon in minerals such as majorite garnet at upper-mantle pressures is therefore of crucial importance to an understanding of the properties of this region of the Earth. The silicate pyroxenes are also believed to comprise a significant portion of the upper mantle¹, and to provide a host for large cations such as sodium and calcium. It has long been assumed that silicon is restricted to the tetrahedral sites within the pyroxene structure. Here, however, we report the high-pressure synthesis and characterization of single crystals of a pyroxene of nominal stoichiometry $\text{Na}(\text{Mg}_{0.5}\text{Si}_{0.5})\text{Si}_2\text{O}_6$, which is the first pyroxene known to contain both tetrahedrally and octahedrally coordinated silicon.

The synthesis of $\text{Na}(\text{Mg}_{0.5}\text{Si}_{0.5})\text{Si}_2\text{O}_6$ pyroxene and its unit-cell parameters based upon X-ray powder diffraction were reported by Gasparik². Single crystals were synthesized from a

stoichiometric mix of high-purity SiO_2 , MgO and $\text{Na}_2\text{Si}_2\text{O}_5$ at 1,600 °C and 150 kbar using a split-sphere anvil apparatus (USSA-2000). The run products consisted of more than 90% by volume of a silicon-rich pyroxene, together with a small quantity of stishovite (SiO_2) and glass. Electron microprobe analysis of the pyroxene showed it to be homogeneous with a composition of 98–99% $\text{Na}(\text{Mg}_{0.5}\text{Si}_{0.5})\text{Si}_2\text{O}_6$ plus 1–2 mol% of an enstatite component ($\text{Mg}_2\text{Si}_2\text{O}_6$). As we were unable to detect any evidence of exsolution of clinoenstatite either optically or by X-ray diffraction we conclude that this component exists in solid solution.

A single crystal of the pyroxene with approximate dimensions $100 \times 50 \times 25 \mu\text{m}$ was used for X-ray diffraction. Unit-cell parameters (Table 1) were determined from the positions of 20 centred reflections in the range $50^\circ < 2\theta < 60^\circ$ with Mo $K\alpha$ radiation. They are consistent with those refined from powder diffraction data², and indicate a monoclinic unit cell similar to that of jadeite³. A hemisphere of intensity data was collected out to $\sin \theta/\lambda = 0.7$, corresponding to two asymmetric units of reciprocal space. Reflections of class $h0l$ with $h + l = 2n + 1$ were the only systematic absences, indicating that the space group of this pyroxene is either $P2_1/n$ or Pn ; intensity statistics indicated a centrosymmetric space group, and thus $P2_1/n$. The 2,408 symmetry-allowed reflections that were measured were reduced to structure factors and averaged in Laue group $2/m$ with $R_{\text{int}} = 2.4\%$ to give 1,137 symmetry-independent reflections, of which 913 with $F > 3\sigma_F$ were included in the refinements. No absorption correction was made as $\mu_{\text{calc}} = 11.2 \text{ cm}^{-1}$.

Least-squares refinements were carried out with RFINE4⁴. The atomic coordinates of an omphacite⁵ refined in space group $P2_1/n$ were used as a starting model, with Na assigned to both the M2 and M2(1) sites, Mg to M1, and Si to M1(1) and the tetrahedral sites. Complex scattering factors for neutral atoms were taken from International Tables⁶. Initial refinements carried out with these site assignments converged smoothly to $R_u = \sum ||F_o| - |F_c|| / \sum |F_o| = 0.045$, $R_w = \sum \sigma_F^{-2} ||F_o| - |F_c||^2 / \sum \sigma_F^{-2} |F_o|^2 = 0.039$, goodness of fit = 1.55 with isotropic temperature factors, and to $R_u = 0.036$, $R_w = 0.036$, goodness of fit = 1.34 with anisotropic thermal parameters for all atoms. Attempts to refine the site distribution subject to the chemistry indicated by the probe analysis (that is, with 2 mol% $\text{Mg}_2\text{Si}_2\text{O}_6$) did not result in any significant deviation from the fixed unit site occupancies used in the final refinement reported here.

The pyroxene structure consists of single tetrahedral chains that cross-link bands of octahedrally coordinated cation sites (Fig. 1). Between these bands are eight-coordinated sites which normally accommodate larger cations such as Ca and Na in the monoclinic pyroxenes. End-member sodium pyroxenes, $\text{NaM}^{3+}\text{Si}_2\text{O}_6$, crystallize in space group $C2/c$, in which all of

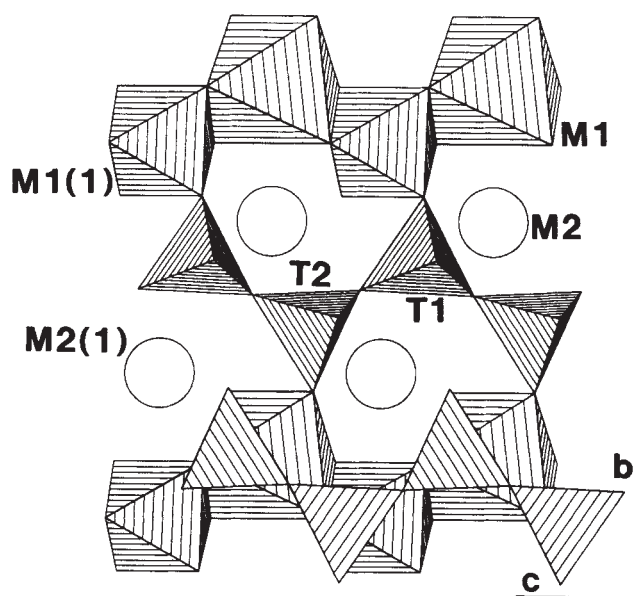


Fig. 1 A polyhedral representation¹⁴ of a portion of the structure of $\text{Na}(\text{Mg}_{0.5}\text{Si}_{0.5})\text{Si}_2\text{O}_6$ clinopyroxene viewed approximately normal to (100).

the octahedral sites (M1) are symmetrically equivalent and all of the Na sites (M2) are also equivalent to one another. The major novel features of the structure of $\text{Na}(\text{Mg}_{0.5}\text{Si}_{0.5})\text{Si}_2\text{O}_6$ pyroxene are due to the presence of silicon in octahedral coordination. The differences in size and charge between Si and Mg result in ordering of these two cations within the M1 sites, while M2 remains occupied by Na. As in omphacitic pyroxenes, cation ordering causes a reduction of the symmetry of the structure from $C2/c$ to $P2_1/n$. In contrast to omphacites, the size difference between the cations in the M1 sites (Mg and Si) is so great as to result in total, rather than partial, ordering, as indicated by the different sizes of the M1 and M1(1) sites. The average M1–O bond distance of 2.100 Å is fairly typical of that found for the Mg octahedra in other pyroxene structures⁷, while the M1(1)–O mean distance of 1.811 Å is similar to that found for octahedrally coordinated silicon in structures such as stishovite⁸, $\text{K}_2\text{Si}_4\text{O}_9$ (ref. 9) and MgSiO_3 ilmenite¹⁰ and perovskite¹¹. However, severe distortions of the coordination polyhedra around all of the other cations are required in order for the pyroxene structure to accommodate such a small octahedral cation as Si. For example, the O1(1)–O1(2) edge which is

Table 1 Crystallographic data for $\text{NaMg}_{0.5}\text{Si}_{0.5}\text{Si}_2\text{O}_6$ pyroxene

	Occ.	x	y	z	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}	B_{eq}
M1	Mg	0.7500	0.6549(2)	0.2500	0.0017(2)	0.0022(2)	0.0041(6)		0.0005(3)		0.56
M1(1)	Si	0.7500	0.8469(2)	0.7500	0.0015(1)	0.0013(2)	0.0036(5)		0.0009(2)		0.40
M2	Na	0.7500	0.0513(2)	0.2500	0.0034(2)	0.0023(3)	0.0060(7)		0.0006(4)		0.83
M2(1)	Na	0.7500	0.4567(3)	0.7500	0.0056(3)	0.0025(3)	0.0091(8)		−0.0011(4)		1.30
T1	Si	0.0447(1)	0.8486(1)	0.2270(2)	0.0012(1)	0.0014(1)	0.0035(3)	−0.0001(1)	0.0006(1)	−0.0002(2)	0.39
T2	Si	0.0372(1)	0.6652(1)	0.7355(2)	0.0012(1)	0.0016(1)	0.0039(3)	−0.0000(1)	0.0010(1)	0.0001(2)	0.40
O1(1)		0.8620(3)	0.8443(3)	0.1017(4)	0.0012(3)	0.0025(3)	0.0038(8)	−0.0001(2)	0.0007(4)	0.0002(4)	0.50
O1(2)		0.8567(3)	0.6934(3)	0.6562(5)	0.0012(3)	0.0021(3)	0.0030(8)	0.0002(2)	0.0001(4)	−0.0001(4)	0.46
O2(1)		0.1234(3)	0.0146(3)	0.3077(5)	0.0020(3)	0.0018(3)	0.0053(9)	−0.0005(2)	0.0010(4)	−0.0004(4)	0.57
O2(2)		0.0982(3)	0.4950(3)	0.7911(5)	0.0022(3)	0.0019(3)	0.0073(9)	0.0005(2)	0.0012(4)	0.0001(4)	0.68
O3(1)		0.1128(3)	0.7665(3)	0.0119(5)	0.0015(3)	0.0026(3)	0.0050(9)	−0.0001(2)	0.0009(4)	−0.0012(4)	0.58
O3(2)		0.0930(3)	0.7527(3)	0.5070(5)	0.0016(3)	0.0025(3)	0.0043(9)	−0.0002(3)	0.0006(4)	0.0007(4)	0.57

Diffraction experiments with Rigaku AFC-5 diffractometer, rotating anode generator, Mo $K\alpha_1$ radiation, graphite monochromator, $\lambda = 0.7093 \text{ Å}$, scan mode $\omega - 2\theta$, ambient temperature and pressure. $a = 9.418(1) \text{ Å}$; $b = 8.647(1) \text{ Å}$; $c = 5.274(1) \text{ Å}$; $\beta = 108.13(1)^\circ$; $V_{\text{cell}} = 408.2(1) \text{ Å}^3$. Space group $P2_1/n$; $Z = 4$; molecular mass = 201.4 $\rho_{\text{calc}} = 3.28 \text{ g cm}^{-3}$.

* Numbers in parentheses are the estimated standard deviation in the last quoted decimal place.

Table 2 Interatomic distances (Å)

M1-O1(1)	2.219 (3) [2]	M1(1)-O1(1)	1.826 (2) [2]
M1-O1(2)	2.090 (2) [2]	M1(1)-O1(2)	1.824 (2) [2]
M1-O2(2)	1.991 (3) [2]	M1(1)-O2(1)	1.782 (3) [2]
Average	2.100	Average	1.811
QE	1.035	QE	1.004
M2-O1(1)	2.333 (3) [2]	M2(1)-O1(2)	2.398 (3) [2]
M2-O2(1)	2.336 (2) [2]	M2(1)-O2(2)	2.434 (3) [2]
M2-O3(2)	2.350 (3) [2]	M2(1)-O3(1)	2.442 (3) [2]
M2-O3(1)	2.677 (3) [2]	M2(1)-O3(2)	2.923 (3) [2]
Average (6)	2.340	Average (6)	2.425
Average (8)	2.424	Average (8)	2.549
QE (6)	1.284	QE (6)	1.417
T1-O1(1)	1.641 (2) [1]	T2-O1(2)	1.637 (2) [1]
T1-O2(1)	1.611 (3) [1]	T2-O2(2)	1.573 (3) [1]
T1-O3(1)	1.629 (2) [1]	T2-O3(1)	1.660 (2) [1]
T1-O3(2)	1.630 (2) [1]	T2-O3(2)	1.642 (2) [1]
Average	1.628	Average	1.628
QE	1.005	QE	1.006

QE is the quadratic elongation as defined by Robinson *et al.*¹³. Figures in square brackets indicate bond multiplicities.

common to both Mg and Si octahedra is only 2.460 Å, compared with a mean O-O edge length around the remainder of the Mg coordination polyhedron of 3.056 Å. This distortion is also reflected in the large quadratic elongation of this octahedron, 1.035, which is comparable with typical values of around 1.01 for Mg in M1 octahedra in other pyroxenes⁷. The sodium coordination polyhedra are also distorted compared to other Na end-member pyroxenes, with the mean M2-O bond being the smallest ever reported, and the mean M2(1)-O distance being longer than that in any of the naturally occurring Na-pyroxenes.

This structure refinement shows that, despite the severe distortions placed on the structure, silicon can be accommodated in substantial amounts in the octahedral sites of clinopyroxenes. There are several possible consequences of this result. The first is that the existence of pyroxenes with a silicon content in excess of 2.00 atoms per six oxygens cannot be excluded on crystal-chemical grounds, especially in those pyroxenes that already contain small cations such as Al and Mg on their M1 sites. Indeed, preliminary experiments indicate significant solubilities of a Na(Mg_{0.5}Si_{0.5})Si₂O₆ component within enstatite, diopside and jadeite at mantle pressures. These substitutions would be expected to change the material properties such as the densities and elasticities of these mantle pyroxenes.

The end-member pyroxene itself was synthesized at 100 and 150 kbar and is likely to be stable to at least 200 kbar. Were the excess of silicon and the presence of sodium the only necessary conditions for the formation of Na(Mg_{0.5}Si_{0.5})Si₂O₆ pyroxene it would therefore replace stishovite as the most likely candidate for the silica-rich phase in the Earth's upper mantle. However, it appears that the most important condition for the stability of this silica-rich sodium pyroxene is the excess of sodium with respect to aluminium, a situation rarely found in nature. An exception may occur at the high-pressure limit of pyroxene stability, where pyroxene transforms to garnet. The incorporation of excess silica could stabilize the pyroxene and expand its stability field to higher pressures. Such behaviour would be analogous to the increase in stability of Ca-Tschermak pyroxene (CaAl₂SiO₆) caused by incorporation of a Ca-Eskola component (Ca_{0.5}AlSi₂O₆) in solution, which incidentally results in a pyroxene with a relative silica excess due to vacancies in the M2 site¹².

The synthesis of single crystals of Na(Mg_{0.5}Si_{0.5})Si₂O₆ pyroxene was performed in the Stony Brook High-Pressure Laboratory which is jointly supported by NSF and the State University of New York, and X-ray diffraction work at the Geophysical Laboratory was supported by NSF and by the Carnegie Institution of Washington.

Received 13 June; accepted 18 July 1988

1. Takahashi, E. & Ito, E. in *High-Pressure Research in Mineral Physics* (eds Manghni, M. H. & Syono, Y.) (American Geophysical Union, Washington DC, 1987).
2. Gasparik, T. *Eos* **69**, 500 (1988).
3. Prewitt, C. T. & Burnham, C. W. *Am. Miner.* **51**, 956-975 (1966).
4. Finger, L. W. & Prince, E. *Natn. Bur. Stand. Tech. Note* 854 (1974).
5. Curtis, L. *et al. Can. Mineralogist* **13**, 62-67 (1975).
6. *International Tables for X-ray Crystallography* Vol. 4 (Kynoch, Birmingham, 1974).
7. Smyth, J. R. & Bish, D. L. *Crystal Structures and Cation Sites of the Rock-Forming Minerals* (Allen and Unwin, Boston, 1988).
8. Sinclair, W. & Ringwood, E. A. *Nature* **272**, 714 (1978).
9. Swanson, D. K. & Prewitt, C. T. *Am. Miner.* **68**, 581-585 (1983).
10. Horiuchi, H., Hirano, M., Ito, E. & Matsui, Y. *Am. Miner.* **67**, 788-793 (1982).
11. Horiuchi, H., Ito, E. & Weidner, D. J. *Am. Miner.* **72**, 357-360 (1987).
12. Gasparik, T. *Geochim. cosmochim. Acta* **48**, 2537-2545 (1984).
13. Robinson, K., Gibbs, G. V. & Ribbe, P. H. *Science* **172**, 567-570 (1971).
14. Fisher, R. X. *J. appl. Crystallogr.* **18**, 258-262 (1984).

²²⁶Ra excesses in mid-ocean-ridge basalts and mantle melting

K. H. Rubin & J. D. Macdougall

Geological Research Division, A-020,
Scripps Institution of Oceanography, La Jolla,
California 92093, USA

The discovery of radioactive disequilibrium among isotopes of the ²³⁸U decay chain in recent volcanic rocks¹⁻⁷ has enhanced our understanding of both the timescale and processes of melting in the mantle. Excesses of ²³⁰Th (half life 75,200 yr) relative to its parent ²³⁴U in some of these rocks indicate that thorium is preferentially partitioned into the melt phase, and constrain the combined duration of melting and magma transport to less than a few hundred thousand years. Recent observations of ²²⁶Ra and ²²⁸Ra excesses over their respective parents ²³⁰Th and ²²⁸Th in individual subaerial volcanoes have led to even more precise determinations of the timescale of magmatic processes^{3,6}. Here we report the observation and implications of ²²⁶Ra (half life 1,620 yr) excesses in mid-ocean-ridge basalts. These excesses are probably produced during melting in the mantle, and the timescale they impose suggests that they are controlled by kinetic processes, rather than chemical equilibrium. Regardless of the process, the ²²⁶Ra excesses constrain the length of the magma transfer times from melting through eruption (and sample collection) to less than about a thousand years.

In the ²³⁸U decay chain each isotope except ²³⁸U and ²⁰⁶Pb is both radiogenic and radioactive. In the state of secular equilibrium, which exists in the absence of chemical fractionation of decay-series elements, the activity of each isotope is the same (activity = $\lambda_i N_i$ where λ_i is the decay constant and N_i the number of atoms of isotope *i*). The timescale for re-attaining secular equilibrium after a chemical disturbance depends on the magnitude of the fractionation but is generally $\ll 10^6$ yr for the entire decay series. Thus it is expected that parent/daughter activity ratios will be unity for all members of the decay series in most of the mantle. It follows that radioactive disequilibrium in mantle-derived rocks must be produced during magmatogenesis and cannot be ascribed to long-term source characteristics⁸. In most cases, current measurement techniques limit the detection of parent-daughter disequilibrium after the fractionation event to a time interval of about five daughter half lives or less.

Each of the mid-ocean-ridge basalt (MORB) samples examined was chosen on the basis of freshness and only glassy rinds were analysed. Table 1 lists sample locations on the East Pacific Rise (EPR) by latitude; details of sample collection are given elsewhere⁹⁻¹². Samples EN113-6D and EN113-7D are both enriched(E)-MORB influenced by the Easter Hot Spot¹², and Vulcan 8-3 is a ferrobalt. The rest of the samples are normal(N)-MORB. None contained more than a per cent or so by volume of phenocrysts (plagioclase and/or olivine). Samples were prepared for analysis following the stringent procedures outlined by Tanzer¹³. Glass chips for analysis were carefully

Table 1 U-Th data

Sample	Lat.	Th p.p.m.	U p.p.m.	(²³⁰ Th) d.p.m. g ⁻¹	(²²⁶ Ra) d.p.m. g ⁻¹	(²³⁴ U/ ²⁴⁸ U)	(²³⁰ Th/ ²³⁸ U)	(²²⁶ Ra/ ²³⁰ Th)	(²³⁸ U/ ²³² Th)	(²³⁰ Th/ ²³² Th)	Th/U wt ratio
A1367-1*†	12.00 N	0.368 ± 0.005	0.143 ± 0.009	0.113 ± 0.006	0.158 ± 0.005	1.01 ± 0.08	1.06 ± 0.08	1.40 ± 0.09	1.20 ± 0.09	1.26 ± 0.07	2.60 ± 0.20
A1370-1B*†	12.83 N	0.166 ± 0.003	0.070 ± 0.004	0.061 ± 0.003	0.086 ± 0.010	1.06 ± 0.08	1.17 ± 0.09	1.41 ± 0.18	1.28 ± 0.11	1.50 ± 0.11	2.40 ± 0.20
A1374-2B*†	12.83 N	0.281 ± 0.005	0.144 ± 0.010	0.116 ± 0.006	0.317 ± 0.008	0.97 ± 0.09	1.08 ± 0.09	2.71 ± 0.06	1.58 ± 0.14	1.70 ± 0.13	1.90 ± 0.10
VUL8-2F‡	21.43 S	0.172 ± 0.003	0.087 ± 0.006	0.067 ± 0.004	0.111 ± 0.003	0.96 ± 0.08	1.04 ± 0.08	1.66 ± 0.11	1.55 ± 0.14	1.61 ± 0.13	2.00 ± 0.10
VUL8-3‡	20.15 S	0.402 ± 0.005	0.216 ± 0.012	0.167 ± 0.009	0.344 ± 0.009	1.03 ± 0.07	1.03 ± 0.07	2.06 ± 0.07	1.65 ± 0.13	1.69 ± 0.13	1.90 ± 0.10
A915-1R1§	20.51 N	0.134 ± 0.004	0.041 ± 0.001	0.039 ± 0.001	0.082 ± 0.010	0.97 ± 0.08	1.30 ± 0.06	2.10 ± 0.26	0.94 ± 0.04	1.22 ± 0.05	3.30 ± 0.20
EN113-4D	27.53 S	0.151 ± 0.002	0.082 ± 0.006	0.064 ± 0.003	0.103 ± 0.004	0.96 ± 0.09	1.04 ± 0.08	1.60 ± 0.10	1.67 ± 0.15	1.73 ± 0.12	1.80 ± 0.10
EN113-6D	26.81 S	1.360 ± 0.020	0.470 ± 0.030	0.360 ± 0.020	0.389 ± 0.004	1.04 ± 0.06	1.02 ± 0.07	1.08 ± 0.02	1.06 ± 0.07	1.09 ± 0.06	2.90 ± 0.20
EN113-7D	26.46 S	1.270 ± 0.020	0.400 ± 0.030	0.360 ± 0.020	0.490 ± 0.010	0.97 ± 0.09	1.21 ± 0.01	1.36 ± 0.08	0.96 ± 0.09	1.16 ± 0.07	3.20 ± 0.20
EN113-13D	24.84 S	0.177 ± 0.002	0.158 ± 0.008	0.111 ± 0.005	0.187 ± 0.006	0.94 ± 0.06	0.94 ± 0.06	1.68 ± 0.09	2.75 ± 0.19	2.57 ± 0.15	1.12 ± 0.06
EN113-36D	22.68 S	0.178 ± 0.006	0.110 ± 0.010	0.072 ± 0.008	0.099 ± 0.008	0.88 ± 0.16	0.91 ± 0.15	1.40 ± 0.20	1.85 ± 0.35	1.68 ± 0.29	1.70 ± 0.20

U and Th isotopes were measured by alpha-counting of the electroplated metals with silicon surface-barrier detectors. ²²⁶Ra was measured by a ²²²Rn-emanation method^{35,36} using average values for 2-4 runs. Reported errors are 2σ statistical counting variance, except where replicate samples (for some ²²⁶Ra measurements) yielded a higher standard deviation. In those cases, the reported error is the standard deviation.

* ALVIN collected samples (all others are from dredge hauls).

† Sample collection discussed in ref. 9.

‡ Sample collection discussed in ref. 10.

§ Sample collection discussed in ref. 11.

|| Sample collection discussed in ref. 12.

selected to avoid any sign of alteration or Mn-oxide coating because secondary phases have quite different U-series characteristics to those of the basalt itself¹⁴. The selected chips were leached in dilute HCl before dissolution.

In the following sections, the convention of denoting activities with parentheses is adopted. Table 1 and Fig. 1 show ²²⁶Ra excesses to be an ubiquitous feature of the analysed basalts. (²²⁶Ra/²³⁰Th) ranges from 1.40 ± 0.09 to 2.71 ± 0.06 and averages 1.78 ± 0.44 for the 8 N-MORB and ferrobassalt V8-3. The two E-MORB show slightly smaller ²²⁶Ra excesses, possibly as a result of decay of an original signature during subcrustal flow of magma from the off-axis hot spot¹². Ra enrichment over Th during MORB genesis might be expected from the fact that Ba, an Ra analogue, is apparently more incompatible than Th for processes occurring in the MORB source¹⁵, but such enrichments have not been reported previously. Values of ²³⁰Th/²³⁸U are lower in these samples than (²²⁶Ra/²³⁰Th), ranging from 0.91 ± 0.2 to 1.30 ± 0.06, with an average of 1.1 ± 0.1 for all 11 samples. Although ²³⁰Th excesses of 25-35% are common in MORB^{4,7}, most of the samples analysed in this study have smaller excesses (with half of them exhibiting equilibrium with ²³⁸U within 2σ error limits). Th/U weight ratios in these samples average 2.3 ± 0.7 and three samples are derived from sources with Th/U ratios < 2 based on their (²³⁰Th/²³²Th).

These data raise the questions of whether radium excesses are primary signatures and if so, of which processes are capable of causing such large enrichments of one highly incompatible element over another. The stringent sample-selection criteria and cleaning procedures, and an understanding of the chemical behaviour of the elements involved, both suggest that the ²²⁶Ra excesses are indeed primary features. For example, whereas U and Ra are very soluble in sea water, Th is not, and Th addition through Mn-oxide coatings or glass hydration products such as palagonite may give spurious ratios in altered samples¹⁴. Ratios involving ²³⁰Th would be especially affected because of the very high (²³⁰Th/²³²Th) of sea water^{16,17}, but those of other nuclides may also be influenced. U, Th and Ra are highly enriched in very slowly growing Mn nodules in the order Th > Ra > U (ref. 18) and exhibit up to 50% ²²⁶Ra deficits¹⁸. It follows that Ra deficits may also occur in the faster-growing Mn-oxide crusts sometimes found even on very young MORB, but they may be smaller if they are limited in part by the supply of ²³⁰Th. At any rate, it is likely that (²²⁶Ra/²³⁰Th) in these Mn coatings does not exceed unity. Thus, if small amounts were inadvertently incorporated into a basalt glass sample, the result would be to decrease, rather than increase, any existing ²²⁶Ra excess relative to ²³⁰Th.

During low-temperature alteration, U (and possibly also Th

and Ra) is added to MORB¹⁹. But the higher solubility of Ra over Th, coupled with evidence for large Th additions to palagonite⁷, might be expected to lead to overall ²²⁶Ra deficits. A deficit was indeed observed for one altered Galapagos Spreading Centre basalt¹⁴, for which (²²⁶Ra/²³⁰Th) = 0.36 ± 0.04. Thus, neither low-temperature alteration processes nor oxide coatings

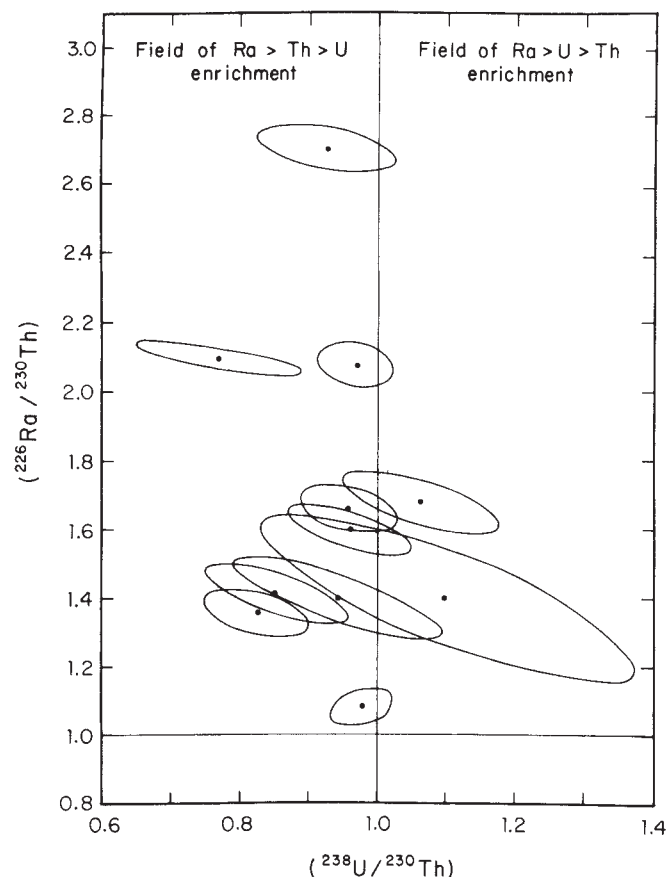


Fig. 1 Values of (²²⁶Ra/²³⁰Th) versus (²³⁸U/²³⁰Th) for samples analysed in this study. Solid lines at (²³⁸U/²³⁰Th) = 1 and (²²⁶Ra/²³⁰Th) = 1 represent secular equilibrium for those isotopes. ²²⁶Ra excesses over ²³⁰Th exist in all 11 samples and are larger than disequilibrium between ²³⁰Th and ²³⁸U. This suggests the incompatibility sequence Ra > Th ≥ U for MORB genesis. Minor and major axes of the error ellipses are determined by the correlated 2σ error in 1/(²³⁰Th) and (²²⁶Ra/²³⁸U) respectively.

are expected to have contributed to ^{226}Ra excesses in our samples.

Another process affecting rock chemistry after emplacement is high-temperature alteration. In hydrothermal chimneys on the Juan de Fuca Ridge, $(^{226}\text{Ra}/^{230}\text{Th}) \gg 1$, and concentrations of ^{226}Ra are many orders of magnitude larger than those found in sea water, suggesting that high-temperature alteration of the oceanic crust removes Ra (K. Kim and G. M. McMurtry, manuscript in preparation). Though none of our rocks shows any evidence for direct high-temperature alteration, this process could lower $(^{226}\text{Ra}/^{230}\text{Th})$ indirectly if the magma had interacted with hydrothermally altered wall rocks. But there is no evidence from elemental or other isotope data for these samples to suggest that this has occurred. In any event, it is very unlikely that high-temperature alteration processes could raise $(^{226}\text{Ra}/^{230}\text{Th})$ in affected rocks.

To address the second question, the three processes most likely to affect trace-element signatures in MORB are (1) melting, (2) magma transport and (3) magma chamber processes. Ra, Th and U are all highly magmaphilic with very low solid-liquid distribution coefficients (k_d). Few measured k_d values have been reported for U or Th (refs 20–22) and none for Ra, but they are generally assumed to be <0.005 , based on observed fractionations in rocks^{7,23}. Thus, magmatic processes involving very small batch melt-fractions ($<1\%$), or very large degrees of crystal fractionation ($>97\%$), are required to separate these elements⁸. In fact, recent theoretical analysis of trace-element distributions in MORB imply very small degrees of equilibrium melting ($<2\%$) in their genesis^{23,24}, contradicting more traditional views based on major-element compositions and equilibrium phase diagrams. The latter suggest melt fractions of 20–30% for MOR tholeiites²⁵. This apparent discrepancy can be resolved if trace- and major-element signatures are developed by melting in different parts of the mantle. Trace-element concentrations would then be controlled by the amount of the trace-element-enriched component in the final melt, and trace-element ratios would be determined by the degree of melting and compositional variability in a region separate from the major-element source^{26,27}.

Closed-system magma chambers can be effectively ruled out as a site for production of ^{226}Ra or ^{230}Th excesses because of the extremely large degrees of crystal fractionation required⁸. Similarly, open-system magma chambers²⁸ are unlikely to produce fractionations of this type as they would require 10^3 to 10^4 eruption/re-injection cycles²⁹, and inferred eruption intervals on the EPR are typically less than about 10^2 yr (ref. 30). Simple batch melting, as mentioned previously, is also unable to produce fractionation of the kind we observe because of the extremely small melt fractions required and the difficulty of transporting and accumulating such small melt volumes on the short timescales permitted.

Recently developed views of melting in the mantle, such as constant-porosity dynamic melt models (refs 23 and 31; R. Williams and J. Gill, manuscript in preparation), achieve chemical fractionation equivalent to that of batch melting for significantly larger melt fractions because the matrix interacts with the melt as an ion-exchanger while the melt moves through it. Significantly, models of this type require continual chemical equilibrium between solid and melt, and instantaneous extraction of melt once the porosity exceeds a given value. The remaining melt plus matrix must remain a system closed to chemical exchange, thus maintaining radioactive equilibrium. Continual redistribution of various nuclides between solid and melt according to their distribution coefficients allows the melt to be continually enriched in the more incompatible elements. The chemical characteristics of the resulting melt are essentially the same as those produced in an extremely small batch melt fraction, but allow for melt volumes as large as 1–2%. In fact, for a judicious choice of distribution coefficients (<0.005 for U and Th, and 0.0001 for Ra (D. McKenzie, personal communication)),

a 1% melt of this type would have $(^{230}\text{Th}/^{238}\text{U}) = 1.25$ and $(^{226}\text{Ra}/^{230}\text{Th}) = 1.5$. It is unlikely, however, that even these models could produce the highest $(^{226}\text{Ra}/^{230}\text{Th})$ of 2–3 observed in these rocks.

In addition, chemical equilibrium, which may not always be established during mantle melting, is a prerequisite of these models. In equilibrium, diffusion of all species must be rapid enough to allow communication between entire solid grains and the melt. For radioisotopes the situation is further complicated by the fact that decay must not alter concentrations significantly during diffusion, which requires that the timescale for diffusion of a radioisotope must be less than, or equal to, its decay lifetime (that is, $t_{\text{diff}} \leq t_{\text{rad}}$)²³. It is clear, however, that for ^{226}Ra the diffusion timescale may be substantially longer. Using the diffusivity, D , measured for U in diopside at $1,240^\circ\text{C}$ ($\sim 10^{-16} \text{ m}^2 \text{ s}^{-1}$ (ref. 21)) to approximate that of Ra, and given that $t_{\text{diff}} \approx a^2/D$ (where a is the grain radius), it can be shown that diffusion-controlled chemical equilibrium cannot be achieved for ^{226}Ra in grains with radii greater than ~ 2 mm. This is smaller than the grain size found in many mantle xenoliths, implying that chemical equilibrium may not be attained for Ra during mantle melting. For these reasons, it is probable that kinetic processes have a function in determining the trace-element characteristics of MORB. These might include, for example, the effects of different diffusion rates for each element in the solid matrix and various geometric factors arising from the disposition of the grains and the diffusing species themselves. In addition, large effects could undoubtedly occur as a result of variable chemical potential gradients caused by different distribution coefficients for each species.

It must be emphasized that although melting is the most likely process for production of radioactive disequilibrium, open-system fluid-phase additions to the melt during transport or storage could also alter the magma's chemistry³². But if this is the case, it is difficult to understand how the fluids become significantly enriched in ^{226}Ra initially, where they originate, and why their occurrence should be ubiquitous over such a vast region of ridge crest.

The observation of ^{226}Ra excess in MORB raises another important issue: the necessity of very rapid melt-transfer to the ridge crest from depths as great as 100 km in the mantle. As all but the most extreme ^{226}Ra excesses would be undetectable after $\sim 8,000$ yr, much shorter transport times are required for reasonable melting conditions. For instance, if the average transport time for melts from their source to the surface is of the order of one ^{226}Ra half life, the average observed 60% excess would require an initial excess of 120% on melting. Regardless of the melting model, transport times and the extent of melting are closely interrelated in terms of their effect on the ^{226}Ra excesses observed in the erupted magmas. Thus, transport times of about 1,000 yr or less are probably required for the excesses observed; for 100 km of melt transport, this implies flow rates of 100 m yr^{-1} . This is many orders of magnitude higher than the rate of upward convection of the mantle at the ridge crests, taken to be about $1\text{--}20 \text{ cm yr}^{-1}$ (ref. 23). Simple one-dimensional porous-flow models cannot accommodate such large melt flow rates, but if flow is focussed along pressure gradients to the ridge crest^{33,34}, rates of $1\text{--}10 \text{ m yr}^{-1}$ may be permissible. Another possibility is that melt movement occurs largely by conduit-like flow analogous to kimberlitic melt transport. In this case, however, arteries carrying melt to the ridge crest would have to be continually replaced as they are disrupted by the (slower) matrix flow.

To summarize, large and variable excesses of ^{226}Ra over ^{230}Th have been found in 11 carefully picked and cleaned MORB glasses from the EPR. Because they occur even in rocks that do not display a ^{230}Th excess, the ^{226}Ra excesses may be an even more sensitive tracer (in very young rocks) of the processes influencing magma chemistry. The results appear to be incompatible with current models for melting in the mantle, which

require continual equilibrium between the evolving melt and the solid matrix. We therefore suggest that kinetic processes are important. The ^{226}Ra excesses also indicate that melt transport rates are considerably higher than predicted by current porous-flow models, and that flow may occur by a different mechanism.

We wish to thank H. Craig and J.-G. Schilling for providing the samples for this work. Helpful comments on the manuscript were provided by S. Galer, D. McKenzie and referees M. Condomines, M. Spiegelman and K. Turekian. This research was supported by NSF.

Received 18 April; accepted 27 July 1988.

1. Oversby, V. M. & Gast, P. W. *Earth planet. Sci. Lett.* **5**, 199–206 (1968).
2. Nishimura, S. *Earth planet. Sci. Lett.* **8**, 293–300 (1970).
3. Capaldi, G., Cortini, M., Gasparini, P. & Pece, R. *J. geophys. Res.* **81**, 350–358 (1976).
4. Condomines, M., Morand, P. & Allegre, C. J. *Earth planet. Sci. Lett.* **55**, 247–256 (1981).
5. Allegre, C. J. & Condomines, M. *Nature* **299**, 21–24 (1982).
6. Bennet, J. T., Krishnaswami, S., Turekian, K. K., Melson, W. G. & Hopson, G. A. *Earth planet. Sci. Lett.* **60**, 61–69 (1982).
7. Newman, S., Finkel, R. C. & Maccougall, J. D. *Earth planet. Sci. Lett.* **65**, 17–33 (1983).
8. Gast, P. W. *Geochim. cosmochim. Acta* **32**, 1057–1086 (1968).
9. Welhan, J. A., Craig, H. & Kim, K. *Trans. Am. geophys. Un.* **65**, 973–974 (1984).
10. Newman, S. thesis, Univ. California (1983).
11. Rise Proj. Group, *Science* **207**, 1421–1433 (1980).
12. Schilling, J.-G., Sigurdsson, H., Davis, A. N. & Hey, R. N. *Nature* **317**, 325–331 (1985).

13. Tanzer, M. O. thesis, Univ. California (1985).
14. Krishnaswami, S., Turekian, K. K. & Bennet, J. T. *Geochim. cosmochim. Acta* **48**, 505–511 (1984).
15. Sun, S.-S., Nesbitt, R. W. & Sharaskin, A. Y. *Earth planet. Sci. Lett.* **44**, 119–138 (1979).
16. Somayajulu, B. L. K. & Goldberg, E. D. *Earth planet. Sci. Lett.* **1**, 102–106 (1966).
17. Krishnaswami, S. *et al. Earth planet. Sci. Lett.* **16**, 84–90 (1972).
18. Moore, W. S. *et al. Earth planet. Sci. Lett.* **52**, 151–171 (1981).
19. Maccougall, J. D. *Earth planet. Sci. Lett.* **35**, 65–70 (1977).
20. Tatsumoto, M. *Earth planet. Sci. Lett.* **38**, 63–87 (1978).
21. Sietz, M. G. *Carnegie Inst. Washington Yb.* **72**, 551–553 (1973).
22. Benjamin, T., Heuser, W. R. & Burnett, D. S. *Proc. Ninth Lunar planet. Sci. Conf.* 1393–1406 (1978).
23. McKenzie, D. *Earth planet. Sci. Lett.* **72**, 149–157 (1985).
24. Richter, F. M. *Earth planet. Sci. Lett.* **77**, 333–344 (1986).
25. Yoder, H. S. *Generation of Basaltic Magma* (National Academy of Sciences, Washington, DC, 1976).
26. Capaldi, G., Cortini, M. & Pece, R. *J. Volcan. geotherm. Res.* **14**, 247–260 (1982).
27. Galer, S. J. G. & O'Nions, R. K. *Terra Cognita* **5**, 273 (1985).
28. O'Hara, M. J. & Mathews, R. E. *J. geol. Soc. Lond.* **138**, 237–277 (1981).
29. Galer, S. J. G. & O'Nions, R. K. *Chem. Geol.* **56**, 45–61 (1986).
30. Macdonald, K. C. A. *Rev. Earth planet. Sci.* **10**, 155–190 (1982).
31. Langmuir, C. H., Bender, J. F., Bence, A. E., Hanson, G. N. & Taylor, S. R. *Earth planet. Sci. Lett.* **36**, 133–156 (1977).
32. Cortini, M. in *Uranium Geochemistry, Mineralogy, Geology, Exploration and Resources* (eds DeVivo, B., Ippolito, F., Capaldi, G. & Simpson, P. P.) 5–11 (Inst of Min. and Metal., London, 1984).
33. Spiegelman, M. & McKenzie, D. *Earth planet. Sci. Lett.* **83**, 137–152 (1987).
34. Morgan, J. P. *Geophys. Res. Lett.* **14**, 1238–1241 (1987).
35. Lucas, H. F. *Rev. Sci. Instrum.* **28**, 680–683 (1957).
36. Broecker, W. S. in *Symp. Diffusion in Oceans and Fresh Waters* (ed. Ichies, T.) 116–145 (Lamont Doherty Geological Observatory, Palisades, New York, 1965).

Cell lineage analysis reveals multipotency of some avian neural crest cells

Marianne Bronner-Fraser* & Scott E. Fraser†

Developmental Biology Center, * Department of Developmental and Cell Biology and † Department of Physiology and Biophysics, University of California, Irvine, California 92717, USA

A major question in developmental biology is how precursor cells give rise to diverse sets of differentiated cell types. In most systems, it remains unclear whether the precursors can form many or all cell types (multipotent or totipotent), or only a single cell type (predetermined). The question of cell lineage is central to the neural crest because it gives rise to numerous and diverse derivatives including peripheral neurons, glial and Schwann cells, pigment cells, and cartilage. Although the sets of derivatives arising from different populations of neural crest cells have been well-documented^{1,2}, relatively little is known about the developmental potentials of individual neural crest cells. We have iontophoretically microinjected the vital dye, lysinated rhodamine dextran (LRD)³ into individual dorsal neural tube cells to mark unambiguously their descendants. Many of the resulting labelled clones consisted of multiple cell types, as judged by both their location and morphology. Cells as diverse as sensory neurons, presumptive pigment cells, ganglionic supportive cells, adrenomedullary cells and neural tube cells were found within individual clones. Our results indicate that at least some neural crest cells are multipotent before their departure from the neural tube.

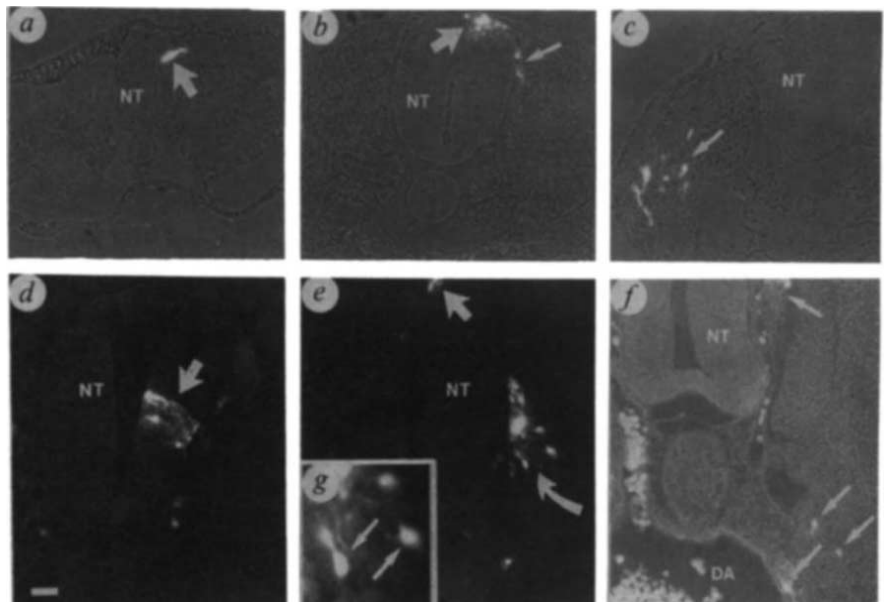
Previous experiments on the neural crest have offered conflicting data on the developmental potential of a single cell. Individual neural crest cells cloned *in vitro* give rise to both pigmented and adrenergic progeny, suggesting they are at least bipotent^{4,5} whereas experiments with monoclonal antibodies which recognize only subpopulations of neural crest cells suggest they are heterogeneous^{6,7}. However, the behaviour of the cells *in vitro* may not reflect their behaviour *in vivo*, and antigens on individual cells may not remain constant. Clonal analysis in the embryo is required to define individual neural crest cell potential. Recent approaches for *in situ* studies of cell lineage have employed either microinjection of large, inert lineage tracer dyes or infection by a recombinant retrovirus^{3,8–15} that introduces a marker gene into the genome^{11–14}. Because the infection is a

random process, identification of a 'clone' relies upon the progeny remaining coherent and is, therefore, not suitable for neural crest cells which disperse during development. Lineage tracers³ provide a direct means of marking a single cell and its derivatives, even if they disperse⁸. Recently, this technique has been used on neuroblasts in the amphibian retina¹⁵, so fragility of vertebrate neuroblasts and dilution of tracer are not insurmountable problems.

We have iontophoretically injected LRD into individual cells and identified their progeny up to two days later. Because LRD is large and membrane-impermeant, it is passed from an injected cell exclusively to its progeny. A single cell in the dorsal aspect of the neural tube (typically between somites number 8 and 28) of stage 10–17 chick embryos was impaled with a LRD-filled microelectrode using an epifluorescence microscope. After dye injection, the presence of a single dye-filled cell was clearly visible by observation with the epifluorescence microscope because the dye spread throughout the cytoplasm. In 86% of the cases ($n = 117$), a single dye-filled columnar epithelial cell was seen spanning the neural tube from its lumen to its outer border. In 8% of the fills, the single labelled cell seemed to extend lamellapods out of the neural tube, as if in the process of emigration. Occasionally (6%) a mitotic pair, apparently connected by a small cytoplasmic bridge, was filled with dye. Histological sections through nine embryos fixed shortly after injection, confirmed that we reliably filled single cells or mitotic pairs of cells in the dorsal neural tube (Fig. 1a). Six of nine fills were of single cells in the dorsal neural tube, one of which had processes extending out of the neural tube. Two of the dye injections filled mitotic pairs of cells. In one case, the filled cell seemed to be dying. Neural tube fills were easily distinguishable from dye fills of overlying ectodermal cells.

In embryos fixed one day after injection (corresponding to the stage of active neural crest migration¹⁶), the descendants of the injected cells appeared viable and were able to migrate away from the neural tube (Fig. 1b, trunk, $n = 12$; Fig. 1c, rhombencephalic, $n = 1$). In 3 embryos, no LRD-labelled cells could be identified, suggesting that the injected cell had died. In the 10 remaining embryos, a variable number of progeny were observed with the fluorescent dextran either uniformly filling the entire cytoplasm (as immediately after injection) or appearing as bright punctuate spots within the cytoplasm. The number of labelled progeny ranged from 4 (trunk level) to approximately 70 (rhombencephalic level, adjacent to the first somite, Fig. 1c). The large number of clearly labelled descendants indicates that

Fig. 1 Photomicrographs of chick embryos fixed after injection of LRD into a single dorsal neural tube cell. *a*, Superimposed bright-field and fluorescence micrograph of an embryo fixed immediately after injection, showing a LRD-labelled neuroepithelial cell which spans the dorsal neural tube (broad arrow). *b*, Superimposed bright-field and fluorescence micrograph of an embryo fixed one day after injection of LRD into a trunk neural tube cell (somites 8 to 28). Some labelled cells remain in the neural tube (broad arrow), whereas others have migrated away (thin arrow). *c*, Superimposed bright-field and fluorescence micrograph of an embryo fixed one day after injection of LRD into a cell in the rhombencephalic neural tube (level of first somite). During the 24-h period, the labelled cell gave rise to approximately 70 progeny (thin arrow), suggesting that the injected cell had divided about 6 times; these were distributed in the cranial mesenchyme over 120 μm in rostrocaudal extent. *d*, Fluorescence photomicrograph of an embryo fixed two days after injection of LRD into a trunk-level dorsal neural tube cell. Labelled cells were only detected within the neural tube (thick arrow). The thickened regions of the individual cells appear to be the cell nuclei which move from the apical to basal surface and back during the cell cycle. *e*, Fluorescence photomicrograph of an embryo fixed two days after injection. Labelled cells were observed within the dorsal neural tube (thick arrow) and in the dorsal root ganglion (curved arrow), where many of the cells appeared neuronal, with large cell bodies and neurites. *f*, Superimposed bright-field and fluorescence micrograph of another section of the same embryo pictured in (*e*). In this section, LRD-labelled cells were observed both in the DRG (thin arrow) and around the dorsal aorta (thin arrow), where adrenergic neural crest cells localize. *g*, Higher power fluorescence micrograph of a portion of the DRG pictured in *e*, showing neurites extending from large cell bodies. The bright cells within the dorsal aorta and in other parts of the sections not marked with arrows are blood cells which autofluoresce brightly. NT, neural tube; DA, dorsal aorta. Scale bar: 25 μm in *a*, *b*, *e*; 30 μm in *c*, *d*, *f*; 10 μm in *g*.



The thickened regions of the individual cells appear to be the cell nuclei which move from the apical to basal surface and back during the cell cycle. *e*, Fluorescence photomicrograph of an embryo fixed two days after injection. Labelled cells were observed within the dorsal neural tube (thick arrow) and in the dorsal root ganglion (curved arrow), where many of the cells appeared neuronal, with large cell bodies and neurites. *f*, Superimposed bright-field and fluorescence micrograph of another section of the same embryo pictured in (*e*). In this section, LRD-labelled cells were observed both in the DRG (thin arrow) and around the dorsal aorta (thin arrow), where adrenergic neural crest cells localize. *g*, Higher power fluorescence micrograph of a portion of the DRG pictured in *e*, showing neurites extending from large cell bodies. The bright cells within the dorsal aorta and in other parts of the sections not marked with arrows are blood cells which autofluoresce brightly. NT, neural tube; DA, dorsal aorta. Scale bar: 25 μm in *a*, *b*, *e*; 30 μm in *c*, *d*, *f*; 10 μm in *g*.

Methods. Embryos ranging in age from the 6- to 24-somite stage were used for injections. A window was cut in the shell of the egg, the vitelline membrane was removed over a short segment of neural tube and a sub-blastodermal injection of India ink was introduced to increase visibility of the embryo. The egg was mounted on the stage of an epifluorescence microscope (Zeiss UEM) and the embryo was viewed with oblique lighting from a fibre optic illuminator. Thin-walled aluminosilicate microelectrodes (resistance 25–40 M Ω) were filled at the tip with LRD (Molecular Probes; 100 mg ml⁻¹), and then back-filled with 1.2 M LiCl. The microelectrode was held by a Huxley-style manipulator (Camden Instruments), and impaled into cells in the dorsal neural tube by briefly 'ringing' the microelectrode with the negative capacitance control of the amplifier. The small but reliable membrane potential recorded through the microelectrode before, during and after the injection was used to determine penetration and to ensure that the electrode had not drifted into a second cell. The membrane potential (range -8 mV to -20 mV) was affected by a tip potential from these dye-filled microelectrodes, but served as a reliable measure of cell impalement. After injection, the eggs were sealed with adhesive tape and returned to the incubator until the time of fixation. Embryos were fixed in 2% or 4% paraformaldehyde in phosphate-buffered saline for 2–4 h, and stored in 70% ethanol, until they were dehydrated, embedded in paraplast, and serially sectioned at 10 μm . The slides were deparaffinized and coverslipped in buffered glycerol (pH 7.8) before observation.

dilution of the injected dye by mitotic activity is not limiting. Interestingly, in the trunk the labelled cells were observed bilaterally in about half the cases. The migrating labelled cells were found both dorso-laterally underneath the ectoderm (where presumptive pigment cells localize) and ventrally in the sclerotome (where presumptive sensory, sympathetic, and adrenomedullary cells will localize). Some of the embryos fixed at 1 day were stained with the HNK-1 antibody, which selectively recognizes neural crest cells as they migrate among the sclerotomal cells¹⁷. All LRD-labelled cells within the sclerotome had HNK-1 immunoreactivity. No sclerotomal cells contained LRD, so it is unlikely that any of the LRD-labelled cells acquired the dye through phagocytosing fragments from dead LRD-filled neural tube cells. As expected, LRD-labelled neural tube cells were not HNK-1 immunoreactive, and the LRD-labelled prepigment cells located underneath the ectoderm were not HNK-1 immunoreactive, or only weakly so.

The majority of the embryos were allowed to survive for two days after injection (total incubation time of 4–4.5 days) by which time neural crest-derived dorsal root ganglia (DRG) and sympathetic ganglia had formed¹⁸. Labelled cells were identified in 34 embryos, 28 of which received injections into a single neural tube cell per embryo, located between the levels of somites 14–23 (Fig. 1*d–g*). In the remaining six embryos, two neural tube cells had been injected per embryo, separated by a minimum distance of 3 somite lengths. In 14 of 34 embryos fixed 2 days after injection, labelled cells were identified only

in the neural tube (Fig. 1*d*). Either these cells did not have descendants in the neural crest or mitotic activity diluted the dye in all of the neural crest cells beyond the level of detectability. The latter seems unlikely because we sometimes observed quite large clone sizes. Of the remaining 20, 60% of clones giving rise to neural crest cells also contained some labelled cells which remained within the neural tube. The fate of these cells remains uncertain, but their presence may suggest that the neural crest is not a pre-segregated population within the neural tube. With the exception of the cells within the neural tube and some presumptive melanocytes, all LRD-labelled cells were HNK-1 immunoreactive (Fig. 2).

The cell types derived from the LRD-labelled cells were assayed by both their location and morphology. After two days of incubation, the LRD uniformly filled the cytoplasm in some cells; in others, the LRD fluorescence was punctate. Some of the labelled cells within the DRG were identifiable as neurons because of their long neurites (Fig. 3*a, b*). One class of neurons had large cell bodies and seemed bipolar or unipolar (Fig. 3*b*), another had large cell bodies and were multipolar (Fig. 3*a*), whereas a third class of neurons had small cell bodies and long, thin processes. Other labelled cells in the DRG had small cell bodies without processes and were presumed to be undifferentiated cells or supportive cells. The LRD-labelled cells observed in the ventral roots were elongated along the path of the motor axons, like Schwann cells (Fig. 3*c*). Labelled cells underneath the ectoderm seemed to be presumptive melanocytes. The

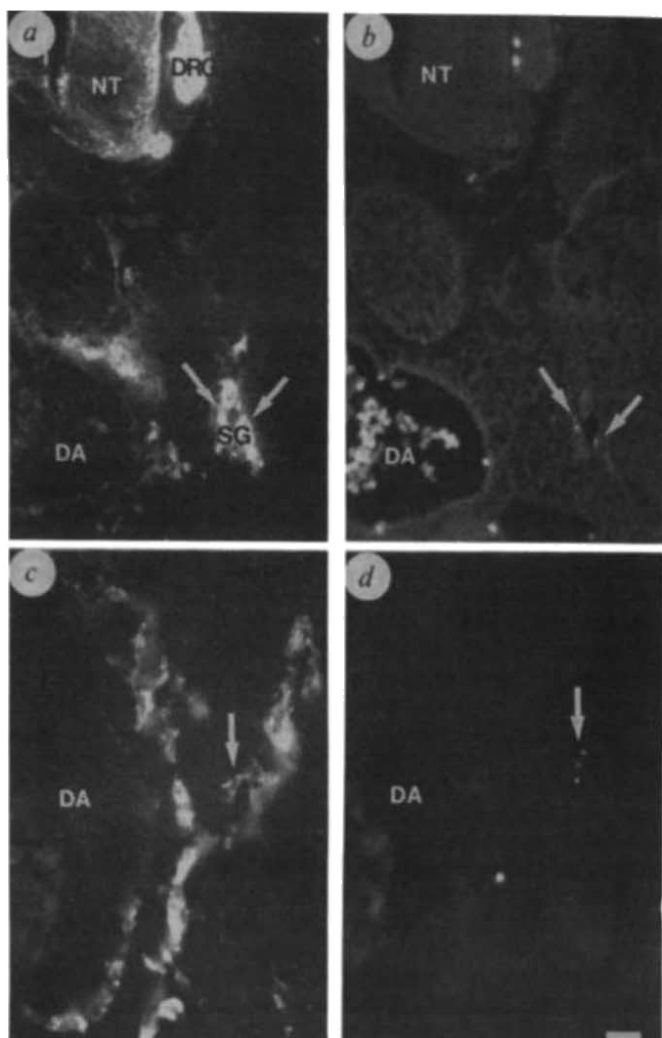


Fig. 2 Fluorescence photomicrographs of transverse sections through embryos fixed two days after injection of LRD into a single dorsal neural tube cell, and stained with the HNK-1 antibody, which recognizes neural crest cells and some of their derivatives¹⁷. A fluorescein isothiocyanate (FITC)-conjugated second antibody was used to label those cells with HNK-1 immunoreactivity. *a*, Epifluorescence micrograph (fluorescein filter set) showing HNK-1 immunoreactivity in DRG and sympathetic ganglia (SG). *b*, Epifluorescence micrograph (rhodamine filter set) of the same section as (*a*) showing LRD-labelled cells within the SG (arrows). This embryo also had LRD-labelled cells in the DRG, around the dorsal aorta, and in the ventral root. *c*, A higher magnification epifluorescence micrograph (fluorescein filter set) of the region around the dorsal aorta of another embryo showing numerous HNK-1 immunoreactive cells in the areas where the neural crest-derived adrenergic cells coalesce to form the adrenal medulla. *d*, Epifluorescence micrograph (rhodamine filter set) of the same section as in *c*, showing that one of the HNK-1 positive cells also contains LRD (arrow). This embryo also had LRD-labelled cells in the dorsal root ganglia and the neural tube. The bright spots in *b* and *d* not marked with arrows were autofluorescent blood cells or debris. Staining with the HNK-1 antibody was performed as described previously¹³. NT, neural tube; DA, dorsal aorta. Scale bar; 30 μ m in *a*, *b*; 15 μ m in *c*, *d*.

labelled cells observed around the dorsal aorta were identified as adrenomedullary cells, based upon both their fibroblast-like morphology and their location (Figs 1*f*, 2, 3*d*).

Table 1 categorizes the variety of labelled derivatives identified in individual embryos by their positions and morphologies. Descendants of single cells were observed in as many as four neural crest derivatives (Fig. 1*e-g*). Sixty per cent of the clones

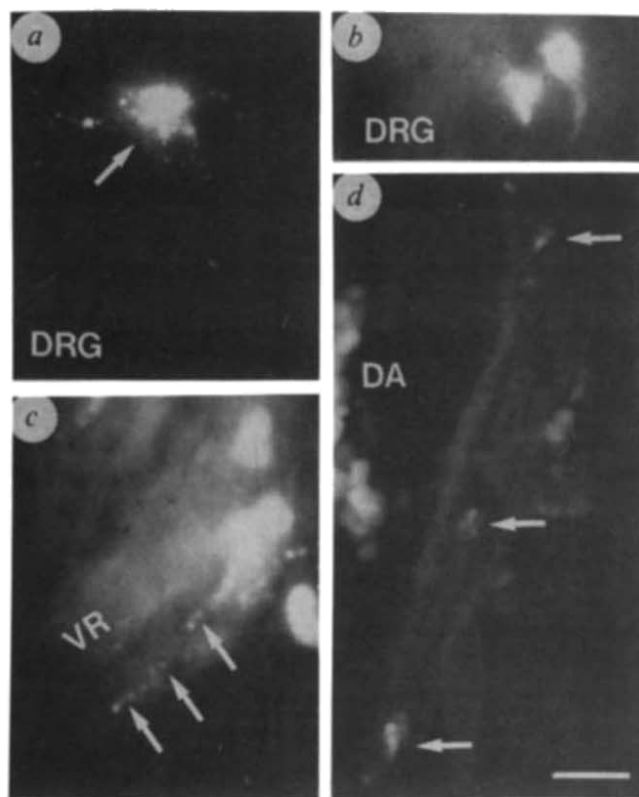


Fig. 3 Fluorescence photomicrographs of recognizable LRD-labelled neural crest derivatives obtained after injection of LRD into a single dorsal neural tube cell. In cells in which the dye fills the cytoplasm, morphological characterization of the LRD-labelled cell was simple. In those cells containing punctate LRD-labelling, it was often possible to assess cell morphology because the punctate fluorescence was distributed throughout the cytoplasm. *a*, *b*, Epifluorescence micrographs of neurite-bearing cells observed in the dorsal root ganglia (DRG) with *a*, punctate and *b*, uniform filling of the cytoplasm by the fluorescent dextran. *c*, Epifluorescence micrograph of the ventral root (VR) shows cells with punctate dextran labelling coursing along the motor axons, a morphology expected of Schwann cells. *d*, Epifluorescence micrograph of the region near the dorsal aorta showing labelled adrenomedullary cells. Images were acquired with an image intensifying video camera (SIT, RCA) and photographed from the screen of the monitor. Thirty-two video frames were averaged with an image processor (Imaging Technology 151) to reduce the noise inherent in the video camera. DRG, dorsal root ganglion; VR, ventral root; DA, dorsal aorta. Scale bar: 10 μ m in *a*; 15 μ m in *b*; 25 μ m in *c*; 30 μ m in *d*.

which gave rise to neural crest cells were distributed in multiple neural crest sites, indicating that the injected precursor was multipotent. The observed multipotency does not seem to depend upon injecting a neural crest founder cell within the neural tube, because even those clones with no members remaining in the neural tube differentiated into multiple cell types (5 out of 8 cases). Interestingly, the number of labelled neural crest cells in different structures was not equivalent, and was generally highest in the DRG. This may indicate that cells injected at the stages and positions we examined contribute predominantly to the DRG. Alternatively, it is possible that the injected cells also gave rise to other structures, but the dye became diluted beyond the limit of detectability. Because our conclusion that some cells are multipotent is based on the positive result of finding lineage tracer in a diverse set of cell types, failure to recognize some progeny due to dye dilution would not affect it.

In a preliminary experiment, we examined the derivatives arising from rhombencephalic neural crest cells. In two embryos

Table 1 Locations of cells derived from clones in embryos fixed 2 days after injection of LRD into the trunk neural tube

No. embryos	NT	DRG	Adrenal	VR	Pigm	SG	Meta
14*	×						
1†		×					
1		×					
1					×		
5	×	×					
1‡			×		×		
1				×	×		
1†	×	×	×				
1†	×	×		×			
1	×	×		×			
2	×	×			×		
1		×		×	×		
1		×	×				×
1†	×	×	×		×		
1		×	×	×		×	
1	×	×	×	×	×		

Injectations were between somite level 14 and 23. An × in a given column signifies the identification of labelled cells in that site. A single neural crest cell was injected per embryo with LRD, except where indicated. NT, neural tube; DRG, dorsal root ganglia; Adrenal, adrenomedullary site; VR, ventral root; Pigm., prepigment cells underneath ectoderm; SG, sympathetic ganglia; Meta, metanephric mesenchyme.

* Two embryos with LRD fills in mitotic pairs of cells; three embryos received two LRD injections separated by 3 somite lengths.

† Embryos in which more than one cell was injected with LRD. These multiple fills were separated by at least 3 segments, and only one of the two filled cells are likely to have survived since the multiple cell types were identified at the same axial level.

‡ This embryo had an LRD fill in a mitotic pair of neural tube cells.

examined two days following injection, labelled cells continued the migration observed at one day after injection (Fig. 1c), and populated normal neural crest sites. In one embryo, labelled cells were observed in the glossopharyngeal ganglion. In the other, labelled cells were observed in the neural tube, the glossopharyngeal ganglion, adjacent to the aorta and in the pharyngeal mesenchyme. This latter case suggests that in cranial regions some neural crest cells are multipotent.

Previous studies examining the sympathoadrenal neural crest lineage¹⁹⁻²⁶ showed that environmental factors may influence the differentiation of neural crest-derived cells. Less is known about the choice between sensory and autonomic lineages. LeLievre and colleagues have found that non-neuronal cells from sensory ganglia can give rise to both sensory and autonomic neurons whereas cells from autonomic ganglia can only give rise to autonomic neurons²⁸, suggesting that a dual sensory/autonomic precursor exists within the neural crest population²⁹. Our results support the notion that some pre-migratory neural crest cells can give rise to both sensory and autonomic lineages. In fact, the majority (60%) of labelled cells which emigrated from the neural tube migrated to multiple neural crest locations and gave rise to morphologically distinct cell types. These results cannot rule out the presence of some predetermined cells within the pre-migratory neural crest, especially in light of the clones (7 out of 20 cases) in which labelled cells were detected only in the sensory ganglia. Because of the possibility of dye dilution, this does not prove that any pre-migratory neural crest cells are predetermined. The ability of single pre-migratory neural crest cells to contribute cells to various embryonic structures poses the question of how these multipotent neural crest precursors are restricted to a given fate. Restriction may be initiated by the emigration of the neural crest cells from the neural tube. Alternatively, restriction may be environment-dependent, mediated by interactions along the migratory pathways or at their final positions. A direct test of

these possibilities is now underway using visual confirmation of the position and morphology of the dextran-labelled cells.

We thank Drs Gabrielle Leblanc, Roberto Perris, Jes Stollberg, Nancy O'Rourke, Rick Wetts and George Serbedzija for helpful comments, Georgia Guillory and Carmen Domingo for technical assistance and Paige Fraser for support in the early phases of this work. This work was supported by grants from NIH, NSF, March of Dimes and a gift from the Monsanto Corporation. S.E.F. is a McKnight Foundation Scholar. M.B.-F. is a Sloan Research Fellow.

Received 25 May; accepted 14 July 1988.

1. LeDouarin, N. M. *Dev Biol.* **30**, 217-222 (1973).
2. LeDouarin, N. M. *The Neural Crest* (Cambridge, University Press, 1982).
3. Gimlich, R. L. & Braun, J. *Dev Biol.* **109**, 509-514 (1986).
4. Sieber-Blum, M. & Cohen, A. M. *Dev Biol.* **80**, 96-106 (1980).
5. Bronner-Fraser, M. E., Sieber-Blum, M. & Cohen, A. M. *J. comp. Neurol.* **193**, 423-434 (1980).
6. Barbu, M., Ziller, C., Rong, R. M. & LeDouarin, N. M. *J. Neurosci.* **6**, 2215-2225 (1986).
7. Girdlestone, J. & Weston, J. A. *Dev Biol.* **109**, 274-287 (1985).
8. Gimlich, R. L. & Cooke, J. *Nature* **306**, 471-473 (1983).
9. Shankland, M. *Dev Biol.* **123**, 97-107 (1987).
10. Cameron, R. A., Hough-Evans, B. R., Britten, R. J. & Davidson, E. H. *Genes Dev.* **1**, 75-84 (1987).
11. Sanes, J. R., Rubenstein, J. L. R. & Nicolas, J.-F. *EMBO J.* **5**, 3133-3142 (1986).
12. Turner, D. L. & Cepko, C. L. *Nature* **328**, 131-136 (1987).
13. Luskin, M. B., Pearlman, A. L. & Sanes, J. R. *Soc. Neurosci. Abstr.* **13**, 183a (1987).
14. Glover, J. C., Gray, C. E. & Sanes, J. R. *Soc. Neurosci. Abstr.* **13**, 183a (1987).
15. Wetts, R. & Fraser, S. E. *Science* **239**, 1122-1145 (1988).
16. Bronner-Fraser, M. *Dev Biol.* **115**, 44-55 (1986).
17. Tucker, G. C., Aoyama, H., Lipinski, M., Turz, T. & Thiery, J. P. *Cell Differ.* **14**, 223-230 (1986).
18. Lallier, T. & Bronner-Fraser, M. *Dev Biol.* **127**, 99-112 (1988).
19. Doupe, A. J., Landis, S. C. & Patterson, P. H. *J. Neurosci.* **5**, 2119-2142 (1985).
20. Doupe, A. J., Patterson, P. H. & Landis, S. C. *J. Neurosci.* **5**, 2143-2160 (1985).
21. Anderson, D. J. & Axel, R. *Cell* **47**, 1079-1090 (1986).
22. Patterson, P. H. A. *Rev. Neurosci.* **1**, 1-17 (1978).
23. Patterson, P. H. & Chun, L. L. Y. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3607-3610 (1974).
24. Landis, S. & Keefe, D. *Dev Biol.* **98**, 349-372 (1983).
25. Leblanc, G. & Landis, S. J. *Neurosci.* **6**, 260-265 (1986).
26. Coulombe, J. N. & Bronner-Fraser, M. *Nature* **324**, 569-572 (1986).
27. Fukada, K. & Patterson, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8795-8799 (1985).
28. LeLievre, C., Schweizer, G., Ziller, C. & LeDouarin, N. *Dev Biol.* **77**, 362-378 (1980).
29. LeDouarin, N. M. *Science* **231**, 1515-1522 (1986).

Calcitonin gene-related peptide acts as a novel vasodilator neurotransmitter in mesenteric resistance vessels of the rat

Hiromu Kawasaki, Koichiro Takasaki, Akira Saito* & Katsutoshi Goto*

Department of Pharmacology, Miyazaki Medical College, 5200 Kiyotake, Miyazaki 889-16, Japan, and

*Department of Pharmacology, Institute of Basic Medical Sciences, University of Tsukuba, Tsukuba, Ibaraki 305, Japan

Systemic blood pressure is controlled by changes in the resistance of the peripheral vascular bed for example in the mesenteric blood vessels¹. The tone of peripheral blood vessels is primarily maintained by sympathetic vasoconstrictor nerves. Although vasodilator innervation has been identified in certain isolated elastic arteries^{2,3}, it is not known whether vasodilator nerves contribute to the regulation of the peripheral resistance vessels. We present pharmacological evidence for the existence of nonadrenergic, noncholinergic (NANC) vasodilator nerves in the mesenteric resistance vessel of the rat and that the resistance is controlled by not only sympathetic vasoconstrictor nerves but also NANC vasodilator nerves. We also show that the neurogenic vasodilation was selectively abolished by depleting endogenous calcitonin gene-related peptide (CGRP), a potent vasodilator neuropeptide^{4,5}, from peripheral nerves. This indicates that CGRP is a novel vasodilator neurotransmitter and may play a role in control of the total peripheral resistance of systemic circulation through a local reflex mechanism.

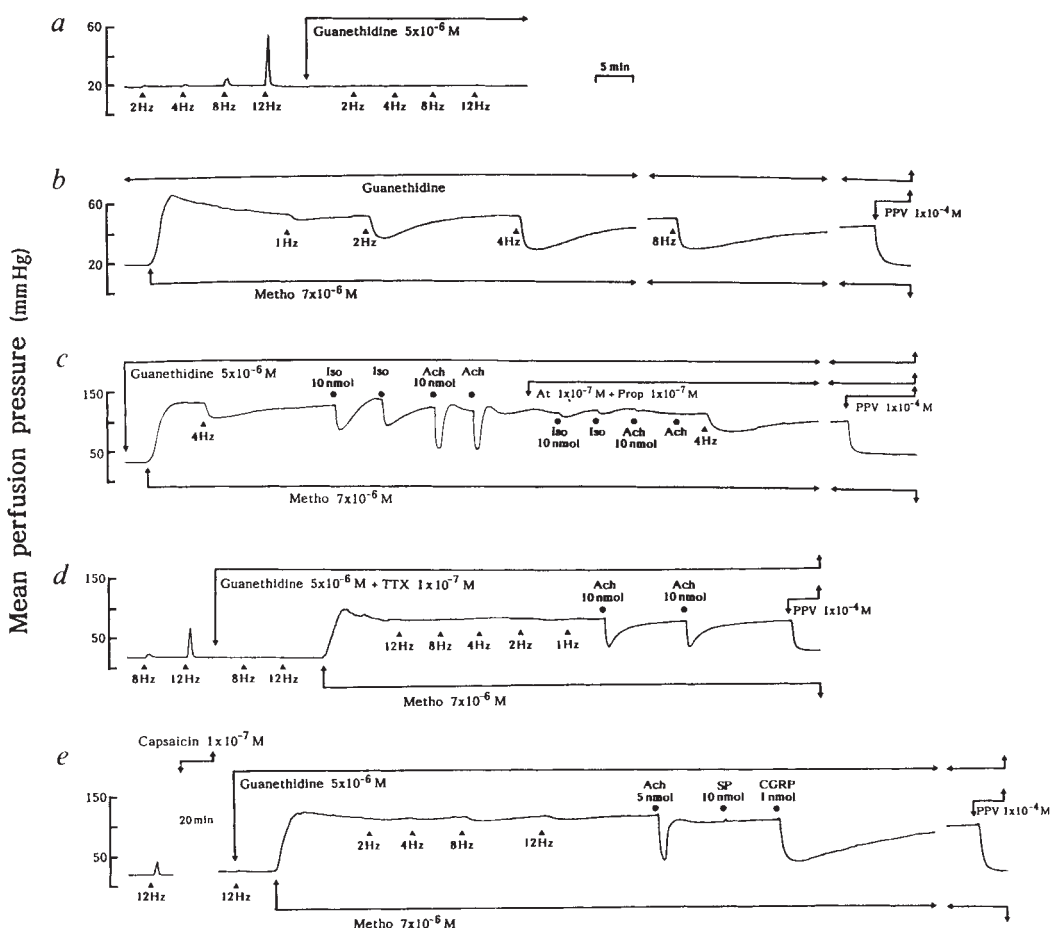
Fig. 1 Typical records of the neurogenic vasodilation induced by periaxillary nerve stimulation (PNS) and effects of various agents in the contracted mesenteric vascular bed of rats. *a*, Frequency-dependent pressure responses to PNS and their abolition by perfusion with guanethidine. *b*, Vasodilation induced by PNS in the preparation contracted by methoxamine. *c*, No effect by atropine (AT) plus propranolol (Prop) on the PNS-induced vasodilation. *d*, Abolition of vasodilation in response to PNS by tetrodotoxin (TTX) or *e*, after treatment with capsaicin. PPV, papaverine. Ach, Acetylcholine. ISO, isoprenaline. SP, substance P. CGRP, calcitonin gene-related peptide. Solid triangles and circles indicate PNS and infusion, respectively.

Methods. The mesenteric vascular beds of male Wistar rats were isolated and prepared for perfusion of Krebs-Ringer bicarbonate solution (Krebs solution) as described previously¹⁷. The preparations were perfused with a modified (see below) Krebs solution at the constant flow rate of 5 ml per min by a peristaltic pump. The preparations were also superfused with the same solution at a rate of 0.5 ml per min to prevent drying. The solution was bubbled with a mixture of 95% O₂/5% CO₂ before passage through a warming coil maintained at 37°C.

Modified Krebs solution of the following composition (mM) was used; NaCl, 120; KCl, 5.0; CaCl₂, 2.4; MgSO₄, 1.2; NaHCO₃, 25; EDTA 2Na, 0.027; and dextrose, 11.0. Changes in the perfusion pressure were measured with a pressure transducer and recorded on a polygraph. After equilibration for 60 min the mesenteric vascular bed was routinely perfused with Krebs solution containing guanethidine (5×10^{-6} M) to block adrenergic neurotransmission and noradrenaline (NA) 7×10^{-7} M to 1×10^{-6} M) or methoxamine (7×10^{-6} M to 1×10^{-5} M) to induce submaximal vasoconstriction that elevated the perfusion pressure from 20 to 50 mm Hg. When the elevated perfusion pressure was lower than 20 mm Hg, the concentration of NA or methoxamine was increased further. In the perfused mesenteric vascular beds, perfusion with prostaglandin F_{2α} up to 1×10^{-5} M caused slight vasoconstriction and perfusion with serotonin (1×10^{-5} M) induced transient contraction. After the elevated perfusion pressure was allowed to stabilize, the preparation was subjected to PNS and infusion of drugs. PNS was for 30 s with an automatic timer through bipolar platinum ring electrodes placed around the superior mesenteric artery. Rectangular pulses 1 ms in duration and of supramaximal voltage (60 V) were applied at 1 to 12 Hz by an electronic stimulator. Ach (5 or 10 nmol); ISO (10 nmol); SP (10 nmol) or CGRP (1 nmol), which were diluted with Krebs solution containing guanethidine and methoxamine, were infused directly into the perfusate proximal to the arterial cannula by an infusion pump for 10 s (solid circles). The volume infused was 0.05 ml. For the chemical sympathectomy of the mesenteric vascular bed, the animal was treated with 6-OHDA dissolved in 0.9% saline containing 0.1% ascorbic acid at 50 mg kg⁻¹ intraperitoneally on days 1 and 2 (24 h apart); the preparation was made on day 3, 24 h after the second administration. The chemical sympathectomy of the preparation was verified by abolition of the pressure response to PNS (16 or 32 Hz) or by disappearance of catecholamine fluorescence by glyoxylic acid method¹⁸. Capsaicin at the concentration of 1×10^{-7} M was perfused for 20 min, and the preparation was rinsed with Krebs solution containing no capsaicin for 60 min. All drugs (except for indometacin and capsaicin which were dissolved in 50% ethyl alcohol) were dissolved in distilled water and diluted in Krebs solution bubbled with a mixture of 95% O₂/5% CO₂ just before being perfused.

Mesenteric vascular beds isolated from male Wistar rats were prepared for perfusion. Periaxillary nerve stimulation (PNS; 2, 4, 8 or 12 Hz) of the perfused mesenteric vascular bed produced a frequency-dependent increase in perfusion pressure, which was abolished by 5×10^{-6} M guanethidine, a sympathetic neuron blocker (Fig. 1a), 1×10^{-7} M tetrodotoxin or 3×10^{-8} M prazosin, an α -adrenoceptor antagonist, or by treating the rats with 6-hydroxydopamine (6-OHDA; 50 mg per kg per day, twice intraperitoneally) a sympathetic neurotoxin. These results indicate that the vasoconstriction was produced by stimulation of periaxillary sympathetic, adrenergic nerves.

In a preparation contracted by continuous perfusion with Krebs solution containing the adrenergic α -receptor stimulant, methoxamine (7×10^{-6} M), and guanethidine (5×10^{-6} M), PNS (1, 2, 4 or 8 Hz) produced a frequency-dependent decrease in perfusion pressure (Fig. 1b). This vasodilation appeared about 20 s after the start of stimulation, reached a maximum at 1 to 2 min after stimulation ended, and the perfusion pressure gradually returned to the pre-stimulation level within 20 to 30 min for the higher frequencies. The PNS-induced vasodila-



tion was not changed by the combined presence of 1×10^{-7} M propranolol, a β -adrenoceptor antagonist, and 1×10^{-7} M atropine, a muscarinic cholinergic antagonist, at the concentration that abolished the vasodilator response to 10 nmol isoprenaline (ISO), a β -adrenoceptor agonist, or 10 nmol acetylcholine (Ach), a cholinergic agonist (Fig. 1c, Table 1). Thus, the vasodilator response to PNS is mediated by NANC nerves. The relaxation induced by PNS in the mesenteric vascular bed contracted by noradrenaline (1×10^{-6} M) was significantly smaller than that in the bed contracted by methoxamine (Table 1).

The PNS-induced vasodilation was not affected by chemical sympathetic denervation with 6-OHDA (50 mg per kg per day, twice intraperitoneally) (Table 1). However, 1×10^{-7} M tetrodotoxin abolished the vasodilator responses to PNS (Fig. 1d, Table 1), consistent with the responses being neurogenic. Indometacin (5×10^{-6} M, a cyclo-oxygenase inhibitor), mepyramine (5×10^{-6} M, a histamine H₁-receptor antagonist), and cimetidine (1×10^{-5} M, a histamine H₂-receptor antagonist) had no effect on the PNS-induced vasodilation (Table 1), suggesting that prostaglandins and histamine are not involved in

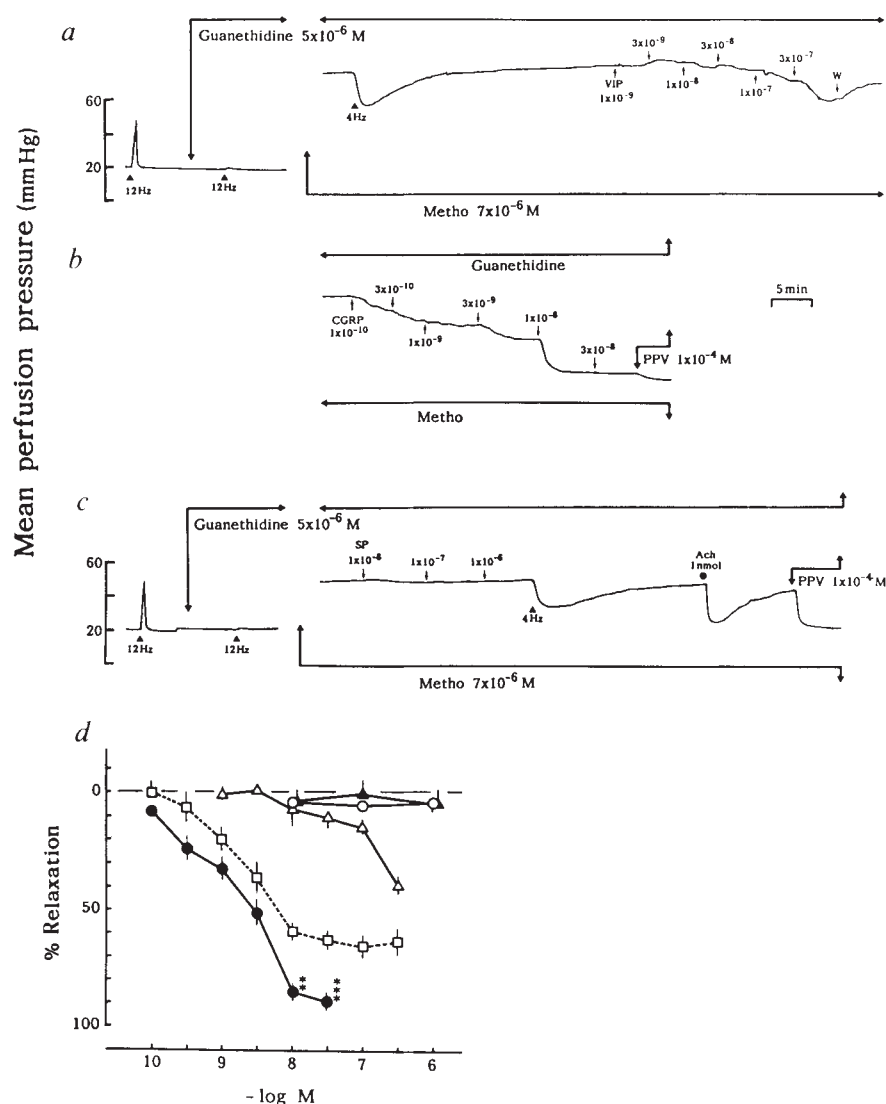


Fig. 2 Typical records of vasodilation induced by vasoactive intestinal polypeptide (VIP, *a*), CGRP (*b*), but not SP (*c*) in perfused mesenteric vascular beds contracted with methoxamine. W, washing with Krebs solution containing no VIP. Panel *d* shows the % relaxation induced by CGRP in the methoxamine-contracted preparation (□, *n* = 8), and by VIP (△, *n* = 4), SP (○, *n* = 4), and neurokinin A (▲, *n* = 4). Vasodilation was expressed as a percentage of the maximum relaxation induced by perfusion with papaverine 1 × 10⁻⁴ M at the end of each experiment. The preparation was perfused with the final concentration of each agent which was achieved by dilution with Krebs solution containing methoxamine and guanethidine. All results are means ± s.e.m. and were analysed statistically by Student's *t* test to compare means of groups. *P* < 0.05 was considered significant. ***P* < 0.01 and ****P* < 0.001 compared with the NA-contracted preparation.

this response. Since tetrodotoxin abolished the response, it is unlikely that electrical stimulus of the rat mesenteric artery releases the endothelium-derived relaxing factor⁶ to produce vasodilation, as reported for the rat-tail artery⁷ and the lung artery of rabbits, cats and monkeys⁸.

Vasoactive neuropeptides such as substance P (SP)⁹, CGRP¹⁰ or vasoactive intestinal polypeptide (VIP)¹¹ have been shown to be present in some blood vessels by immunohistochemical methods. Exogenous SP, CGRP and VIP produce vasodilation in contracted blood vessels^{5,11}. To identify possible neurotransmitters involved in the vasodilation mediated by NANC nerves in the mesenteric vascular bed, we used capsaicin, which depletes SP and CGRP but not VIP from nerves^{12,13}. When the perfused mesenteric vascular bed was treated with 1 × 10⁻⁷ M capsaicin for 20 min, PNS did not cause vasodilation (Fig. 1e, Table 1). These results clearly indicate that the NANC vasodilator response is mediated by the capsaicin-sensitive nerves. The infusion of 5 nmol Ach and 1 nmol CGRP caused marked vasodilation in the preparation treated with or without capsaicin (Fig. 1e), whereas infusion of 10 nmol SP did not bring about relaxation (Fig. 1e). The CGRP- but not the Ach-induced vasodilation was slow in both onset and decay, which mimicked the PNS-induced vasodilation. In the mesenteric vascular bed contracted by methoxamine, perfusion of neither SP (1 × 10⁻⁸ M to 3 × 10⁻⁶ M) nor neurokinin A (1 × 10⁻⁸ M to 1 × 10⁻⁶ M) caused vasodilation (Fig. 2c and d), suggesting that SP-like substances are not involved in the vasodilation. Perfusion of VIP (1 × 10⁻⁹ M to 3 × 10⁻⁷ M) relaxed the contracted mesenteric

Table 1 Effects of various agents and treatments on neurogenic vasodilation

Agents/treatments	Concentration (M)	n	PNS (Hz)	
			4	8
% Relaxation				
Methoxamine-contracted				
Control		9	57.0 ± 4.6	70.3 ± 4.5
Capsaicin	1 × 10 ⁻⁷	9	1.3 ± 2.0*	4.5 ± 3.1*
Tetrodotoxin	1 × 10 ⁻⁷	9	5.9 ± 2.0*	5.8 ± 1.6*
Atropine	1 × 10 ⁻⁷	8	51.9 ± 7.7	60.5 ± 9.4
+Propranolol	1 × 10 ⁻⁷	8	51.9 ± 7.7	60.5 ± 9.4
Indometacin	5 × 10 ⁻⁶	6	62.9 ± 9.5	64.8 ± 5.3
Cimetidine	1 × 10 ⁻⁵	5	46.2 ± 5.3	62.6 ± 4.6
Mepyramine	5 × 10 ⁻⁶	5	51.4 ± 7.2	66.2 ± 8.3
6-Hydroxidopa-mine	50 mg per kg per day, twice	4	49.9 ± 9.4	56.0 ± 8.5
Noradrenaline-contracted				
Control		6	9.5 ± 2.3*	24.0 ± 4.0*

The preparation was routinely perfused with 5 × 10⁻⁶ M guanethidine. Vasodilation was expressed as a percentage of the maximal relaxation induced by perfusion of 1 × 10⁻⁴ M papaverine at the end of each experiment.

* *P* < 0.001, compared with control in the methoxamine-contracted preparation.

vascular bed only at higher concentrations (Fig. 2a and d), but perfusion of CGRP (1 × 10⁻¹⁰ M to 3 × 10⁻⁸ M) produced marked vasodilation (Fig. 2b and d). This is more evidence that the vasodilator substance released by PNS is most likely CGRP.

The role of CGRP as a possible transmitter for mesenteric vasodilation was examined further by immunohistochemistry.

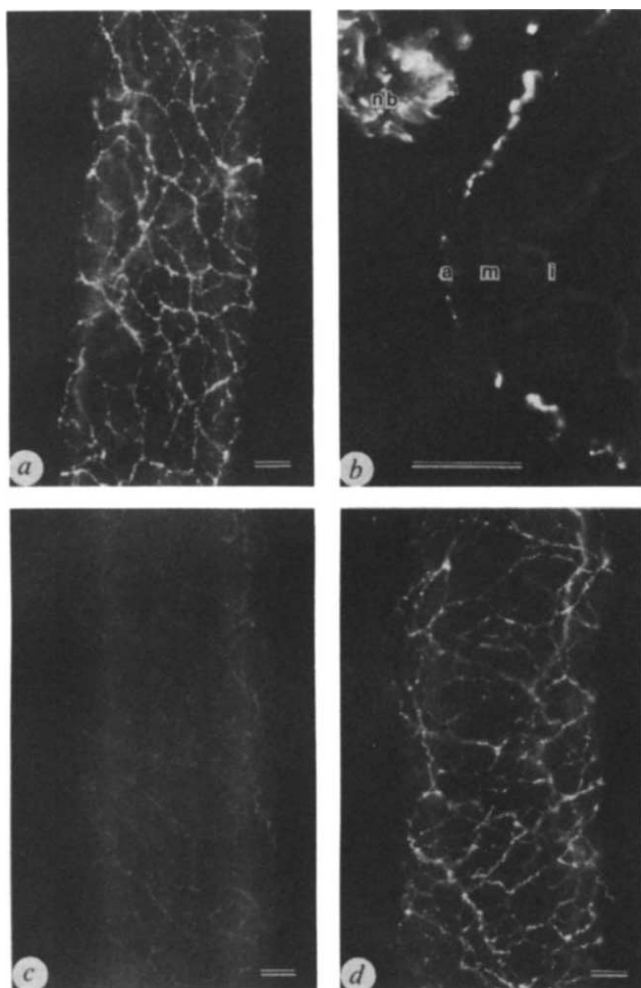


Fig. 3 CGRP-like immunoreactive nerves in the second branch of the mesenteric artery. *a*, There were many CGRP-like immunoreactive nerves in the control tissue. *b*, Immunoreactive material was observed in the adventitial layer (*a*) but not in media (*m*) or intima (*i*). *nb*, nerve bundle. *c*, after exposure of the tissue to capsaicin (3×10^{-7} M) for 20 min, the intensity of CGRP-like immunoreactive nerves reduced markedly. *d*, CGRP-like immunoreactive nerves were observed in the mesenteric artery of rats treated with 6-OHDA. Bar, 50 μ m.

Methods. Mesenteric arteries were fixed in paraformaldehyde and picric acid for 24 to 72 h. The tissues were dehydrated with ethanol, cleared with xylene and rehydrated. Some arteries were frozen and sectioned on a cryostat. Antibodies to CGRP (Cambridge Research, Biochemical) were applied at a 1:1600 dilution for 16 to 24 h at 4 °C. Biotinylated anti-rabbit immunoglobulin G and then fluorescent isothiocyanate-labelled avidin were applied to the tissues. Tissue was mounted in buffered glycerol and examined under a Zeiss fluorescence microscope. The tissue incubated with CGRP antibodies absorbed with CGRP (3×10^{-5} M) was the control.

saicin. These results show that there are NANC vasodilator nerves in the rat mesenteric resistance vessels, and indicate that CGRP is a neurotransmitter of NANC vasodilator nerves.

The PNS-induced vasodilator response in mesenteric vascular beds contracted by noradrenaline was significantly smaller than that in the preparation contracted by methoxamine. The vasodilation induced by CGRP was significantly smaller in the noradrenaline- than in the methoxamine-contracted preparations (Table 1). Preliminary results showed that perfusion with noradrenaline at low concentrations (1×10^{-7} M and 3×10^{-7} M) that produce slight vasoconstriction markedly inhibited the PNS-induced vasodilation in the mesenteric vascular bed contracted by methoxamine. This implies that circulating noradrenaline or neuronally-released noradrenaline from sympathetic adrenergic nerves counteracts the NANC nerve-mediated vasodilation.

Since CGRP-containing nerves are sensitive to the neurotoxic effect of capsaicin, a sensory neurotoxin, CGRP is probably contained in sensory nerves¹⁵. The activation of peripheral terminals of sensory nerves has been shown to cause antidromic vasodilation through the local axon reflex¹⁶. The splanchnic circulation mainly contributes to the total peripheral resistance¹. Thus, the NANC vasodilator nerves in the mesenteric resistance vessels may play an important role in control of the mesenteric vascular tone and the level of arterial blood pressure through a local reflex mechanism.

We thank Mr Keizo Masumoto for excellent technical assistance and Mrs Keiko Kawabata for preparation of the figures.

Received 16 February; accepted 18 July 1988

1. Lundgren, O. *Fedn Proc.* **42**, 1673–1677 (1983).
2. Lee, T. J.-F., Su, C. & Bevan, J. A. *Circulation Res.* **42**, 535–542 (1978).
3. Cohen, R. A., Shepherd, J. T. & Vanhoutte, P. M. *Circulation Res.* **52**, 16–25 (1983).
4. Brain, S. D., Williams, T. J., Tippins, J. R., Morris, H. R. & MacIntyre, I. *Nature* **313**, 54–56 (1985).
5. Marshall, I., Al-Kazwini, S. J., Holman, J. J. & Craig, R. K. *Eur. J. Pharmacol.* **123**, 217–222 (1986).
6. Furchgott, R. F. & Zawadzki, J. V. *Nature* **288**, 373–376 (1980).
7. Kotecha, N. & Neild, T. O. *Clin. exp. Pharmac. Physiol.* **13**, 249–257 (1986).
8. Frank, G. W. & Bevan, J. A. *Am. J. Physiol.* **244**, H793–H798 (1983).
9. Barja, F., Mathison, R. & Huggel, H. *Cell Tiss. Res.* **229**, 411–422 (1983).
10. Saito, A. & Goto, K. *J. Pharmacobiodyn.* **9**, 613–619 (1986).
11. Lee, T. J.-F., Saito, A. & Berezin, I. *Science* **224**, 898–901 (1984).
12. Buck, S. H. & Burks, T. F. *Pharmac. Rev.* **38**, 179–226 (1986).
13. Saito, A., Kimura, S. & Goto, K. *Am. J. Physiol.* **250**, H693–H698 (1986).
14. Lundberg, J. M. *et al. Eur. J. Pharmac.* **108**, 315–319 (1985).
15. Franco-Cereceda, A. *et al. Peptides* **8**, 399–410 (1987).
16. Muramatsu, I. *Jap. J. Pharmac.* **43**, 113–120 (1987).
17. Kawasaki, H. & Takasaki, K. *J. Pharmac. exp. Ther.* **229**, 816–822 (1984).

Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection

Austin L. Hughes & Masatoshi Nei*

Center for Demographic and Population Genetics, the University of Texas Health Science Center at Houston, Houston, Texas 77225, USA

The major histocompatibility complex (MHC) loci are known to be highly polymorphic in humans, mice and certain other mammals, with heterozygosity as high as 80–90% (ref. 1). Four different hypotheses have been proposed to explain this high degree of polymorphism: (1) a high mutation rate², (2) gene conversion or interlocus genetic exchange^{3–5}, (3) over dominant selection^{6,7} and (4) frequency-dependent selection^{8,9}. In an attempt to establish which of these hypotheses is correct, we examined the pattern of nucleotide substitution between polymorphic alleles in the region

* To whom correspondence should be addressed.

Numerous CGRP-like immunoreactive (CGRP-I) fibres were found to be present in the second and third branches of the mesenteric artery (Fig. 3a), but few SP-like or VIP-like immunoreactive nerves were found (data not shown). This suggests that CGRP is unlikely to coexist with SP, although they do in the vagal and spinal ganglia¹⁴. CGRP-like immunoreactivity was markedly diminished after treatment with capsaicin (3×10^{-7} M), but it was not affected by 6-OHDA treatment, indicating that CGRP-I nerves are distinct from sympathetic adrenergic nerves. Admittedly, the NANC vasodilator response disappeared when endogenous CGRP was depleted by cap-

of the antigen recognition site (ARS)^{10,11} and other regions of human and mouse class I MHC genes. The results indicate that in ARS the rate of nonsynonymous (amino acid altering) substitution is significantly higher than that of synonymous substitution in both humans and mice, whereas in other regions the reverse is true. This observation, together with a theoretical study¹² and other considerations, supports the hypothesis of overdominant selection (heterozygote advantage).

The classical class I MHC antigens are glycoproteins expressed on the surface of all nucleated cells. They bind to intracellularly processed foreign proteins, providing the context for recognition of the foreign proteins by cytotoxic T lymphocytes. The extracellular portion of the molecule includes two variable domains (α_1 and α_2) and a third conserved domain (α_3) which associates noncovalently with β_2 -microglobulin. The α_1 , α_2 and α_3 domains are encoded by exons 2, 3 and 4 of the gene respectively. Exon 5 encodes the transmembrane portion of the molecule and exons 6, 7, and 8 encode the cytoplasmic tail^{1,7,13}. Bjorkman and co-workers¹¹ have identified the 57 amino acid residues of the human class I molecule that are involved in the recognition of processed foreign antigens. They are all located in the α_1 and α_2 domains. It is therefore interesting to examine the pattern of nucleotide substitution in the codons responsible for these residues, in comparison with that for other regions of the gene.

If MHC molecules are subject to overdominant selection and ARS is responsible for the selection, the rate of nonsynonymous substitution will be higher than that of synonymous substitution in ARS. This reasoning is based on the theoretical prediction that in the presence of overdominant selection the rate of codon substitution is enhanced compared with that for neutral alleles¹². In addition only nonsynonymous substitutions would be subject to overdominant selection as synonymous substitutions are more or less neutral. The increase in rate of codon substitution under overdominant selection is due to the selective advantage of heterozygotes carrying new mutant alleles over average

individuals. In contrast, if a high mutation rate or interlocus genetic exchange is responsible for MHC polymorphism and no selection is involved, the rates of synonymous and nonsynonymous substitution are expected to be more or less the same. It is therefore possible to distinguish between the overdominance hypothesis and the first two hypotheses by examining the relative rates of synonymous and nonsynonymous substitution. Some form of frequency-dependent selection may also accelerate the rate of nonsynonymous substitution, but this hypothesis seems to have other problems, as will be discussed later.

We collected from the literature 12 sequences for functional alleles at the human class I polymorphic loci *HLA-A*, *B* and *C*. For each of the 66 pairwise comparisons among these sequences, we computed the numbers of nucleotide substitutions per synonymous site (d_s) and per nonsynonymous site (d_N) using Nei and Gojobori's method I^{14,15}. Since the number of pairwise comparisons is large, we present only the mean of d_s and d_N for each of the three intralocus and three interlocus comparisons (Table 1). In ARS, mean d_N is greater than mean d_s for all groups of comparisons, and the difference between the overall means of d_s and d_N is highly significant. Also, d_N was greater than d_s in all 66 pairwise comparisons. We applied the same analysis to the remainder of exons 2 and 3, excluding ARS, and to exon 4; Exons 1 and 5–8 were not analysed because alignment was uncertain and sequence data were not available for all alleles. In non-ARS regions, the relationship of d_N and d_s is opposite to that of ARS, d_s being greater than d_N in most pairwise comparisons (Table 1). The overall mean of d_s is two to ten times greater than that of d_N in these regions.

We compared the same regions of nine class I MHC alleles from the mouse polymorphic loci *H2-K*, *L* and *D*. Here d_N again exceeds d_s in most pairwise comparisons (30 of 36) in ARS (Table 1). The overall mean of d_N for ARS is also considerably higher than the mean d_s although the difference between them is statistically significant only in the intralocus comparison. The mean d_N for ARS is much higher than that for the remainder

Table 1 Mean numbers of nucleotide substitutions per 100 synonymous sites (d_s) and per 100 nonsynonymous sites (d_N)

Locus (No. sequences)	Comparisons (No.)	Antigen recognition site (ARS) ($N = 57$)		Remaining codons in exons 2 & 3 ($N = 124, 125$) [†]		Exon 4 ($N = 92$)		
		d_S	d_N	d_S	d_N	d_S	d_N	
Human								
<i>A</i> (5)	<i>vs A</i> (10)	3.5±2.0	13.3±2.2***	2.5±1.2	1.6±0.5	9.5±3.0	1.6±0.7**	
	<i>vs B</i> (20)	9.1±3.3	25.1±3.4***	11.9±3.0	5.8±0.7*	35.1±8.1	2.2±0.7***	
	<i>vs C</i> (15)	7.1±3.4	21.9±3.5***	17.1±4.0	7.5±1.4*	34.9±7.8	2.1±1.2***	
<i>B</i> (4)	<i>vs B</i> (6)	7.1±3.1	18.1±2.8**	6.9±2.0	2.4±0.7*	1.5±1.1	0.5±0.4	
	<i>vs C</i> (12)	6.0±2.2	22.9±3.4***	14.3±3.2	5.7±1.1*	10.6±4.0	3.1±1.2	
<i>C</i> (3)	<i>vs C</i> (3)	3.8±2.5	8.8±2.2	10.4±2.8	4.8±1.1	2.1±1.5	1.0±0.6	
Overall means								
Intralocus	(19)	4.7±2.6	14.1±2.4***	5.1±2.1	2.4±0.8	5.8±2.0	1.1±0.6**	
Interlocus	(47)	7.7	23.5	14.2	6.3	28.8	2.4	
All comparisons	(66)	6.8±2.3	20.8±2.3***	11.6±2.1	5.2±0.8**	22.1±4.4	2.4±0.7***	
$d_S > d_N: d_N > d_S$		0:66		63:3		61:3‡		
Mouse								
<i>K</i> (4)	<i>vs K</i> (6)	15.0±4.5	22.9±3.3	8.7±2.2	5.8±1.0	2.3±1.3	4.0±1.0	
	<i>vs L</i> (16)	14.9±3.6	21.9±2.5	9.0±1.8	6.8±0.9	2.7±1.5	5.0±1.2	
	<i>vs D^d</i> (4)	14.6±4.7	20.6±2.6	7.0±2.2	6.4±1.1	1.2±0.7	6.1±1.5**	
<i>L</i> (4)	<i>vs L</i> (6)	11.4±4.0	19.5±3.1	8.8±2.3	6.8±1.2	0.0±0.0	2.5±0.8**	
	<i>vs D^d</i> (4)	10.4±4.3	20.5±3.3	5.6±2.1	6.6±2.1	1.5±1.5	4.8±1.3	
Overall means								
Intralocus	(12)	13.2±4.3	21.2±3.2*	8.8±2.3	6.3±1.1	1.2±0.9	3.6±0.9**	
Interlocus	(24)	14.1	21.5	8.1	6.6	2.3	5.3	
All comparisons	(36)	13.8±3.2	21.4±2.4	8.3±1.7	6.5±0.8	1.9±1.0	4.7±0.9**	
$d_S > d_N: d_N > d_S$		6:30		29:7		4:31‡		

The standard errors of mean d_s and d_N were computed by adapting the method of Nei *et al.*³³. (A computer program is available upon request.) The standard errors for interlocus comparisons were not computed because of their complexity. The difference between d_s and d_N is significant at 5% level (*), at 1% level (**), and at 0.1% level (***). Human sequences used: A2³⁴, A3³⁵, A11³⁶, Aw24³⁷, Aw68³⁶, B13³⁸, B27³⁹, B44⁴⁰, Bw58⁴¹, Cw1 and Cw2⁴² and Cw3⁴³. Mouse sequences used: K^d (ref. 44), K^k (ref. 45), K^b (ref. 46), K^{w29.7} (ref. 47), L^d (ref. 48), D^p (ref. 49), D^k and D^b (ref. 50), and D^d (ref. 51). In the mouse, the locus designation for haplotype H-2^d is used (ref. 7); D^p, D^k, and D^b are considered allelic to L^d, not to D^d (ref. 50, 52). N = Number of codons compared. [†] N is 124 for the mouse and 125 for the human. [‡] Excluding ties with $d_s = d_N = 0$.

of exons 2 and 3 or for exon 4. In comparisons of the non-ARS regions of exons 2 and 3, d_s exceeds d_N in the majority of cases. In the case of exon 4, however, d_N exceeds d_s in most pairwise comparisons although this is due to the reduction of d_s rather than the increase of d_N .

It is important to note that in most eukaryotic genes so far studied the number of nucleotide substitutions between polymorphic alleles is very small (0.0001–0.02 per nucleotide site) and that most of these substitutions are synonymous¹⁵. The high values of d_N and d_s observed here are therefore unique to the polymorphic alleles at the MHC loci. Furthermore, our observation that d_N is significantly greater than d_s in a functionally important part of the gene seems to be without precedent.

When the human and mouse genes are compared, d_s and d_N are both greater than those for intraspecific comparisons except in some cases which will be discussed later (Table 2). In ARS, mean d_N is again greater than mean d_s , though the difference is not as large as that for intraspecific comparisons. In the non-ARS regions mean d_s is significantly greater than mean d_N . Therefore, both intraspecific and interspecific comparisons give essentially the same results and support the hypothesis that the rate of nonsynonymous substitution in ARS is enhanced by overdominant selection, whereas the other regions are subject to purifying selection.

In intralocus comparisons of allelic sequences in humans (Table 1), mean d_s is nearly the same for all three regions of the gene, whereas mean d_N is considerably higher in ARS than in the other regions. This result is consistent with the overdominance hypothesis. However, the relationships between d_s and d_N in interlocus comparisons are complicated. Although d_N remains substantially higher in ARS than in the other regions, d_s is generally higher in the non-ARS regions than in ARS. For example, d_s in exon 4 is as high as that for the comparison between human and mouse genes. There are two factors involved in this anomalous feature. One is that the ARS of *HLA* molecules includes a substantial number of highly conserved amino acids such as tyrosine, glycine, glutamic acid, and threonine¹¹ and each of these amino acids is often encoded by the same codon for all alleles. The other is that exon 4 of the alleles from locus *A* varies considerably from that of the alleles from loci *B* and *C* in synonymous sites, although the similarity in nonsynonymous sites is very high. One possible explanation for the latter observation is exon shuffling, which would have exchanged exon 4 of locus *A* with one from another class I locus (not *B* or *C*). Although no donor sequence has been identified to support this explanation, a similar exon shuffling seems to have occurred between alleles at the *A* locus¹⁶.

The mouse data also show some anomalous features. Unlike the human genes, d_s in the mouse is lower in the non-ARS region than in the ARS. Particularly in exon 4, d_s is very low. It is tempting to hypothesize that there has been exchange of exon 4 among the *K*, *L* and *D* loci, but again, there is no evidence to support this.

As mentioned above, the $d_N:d_s$ ratio in ARS is lower in interspecific comparisons than in intraspecific comparisons. This observation probably reflects the fact that some amino acids are conserved even in the ARS and the frequency of parallel and

backward mutations at the amino acid level increases as d_N increases. Table 2 shows that the mean d_s between human and mouse genes is about 0.32. This value is less than half the average d_s for other genes¹⁷. Part of the reason for this low d_s value seems to be the unusually high GC content (70–93%) in the third codon positions of the MHC genes in humans and mice.

Of the four hypotheses for explaining MHC polymorphism, that of high mutation rate can be rejected because the mutational divergence (d_s) between human and mouse MHC genes is not particularly high (see also ref. 7). The hypothesis of gene conversion involving a short DNA segment⁵ has also been disputed⁷. Although interlocus genetic exchange apparently occurs with a certain probability it cannot be the major cause of polymorphism at MHC loci because without selection unequal values of d_N and d_s in ARS and the other regions cannot be explained. The low polymorphism of nonclassical class I genes which comprise a large number of loci (*Qa* and *Tla* loci in the mouse)¹⁸, in contrast to the high polymorphism of the classical class I genes which include only three loci, is also inconsistent with the hypothesis of interlocus genetic exchange.

There are three other lines of evidence supporting the overdominance hypothesis. First, it is known that in both mice⁶ and humans¹⁹, cytotoxic lymphocytes preferentially recognize certain combinations of viral antigen and classical class I antigen. As a particular class I antigen may provide enhanced recognition of a particular pathogen, giving the individual improved resistance to that pathogen, individuals expressing various types of class I antigen may have a selective advantage in a population exposed to a diverse array of pathogens; such individuals will be those which are heterozygous at all or most loci. Second, it has been shown that the level of heterozygosity relative to the number of alleles observed at the *HLA-A*, *B* and *C* loci is too high to be consistent with the neutral expectation^{20,21}. This suggests that the polymorphism at these loci is maintained by some form of balancing selection. Third, the extent of divergence between different alleles at MHC loci is generally very high, and some pairs of polymorphic alleles seem to have persisted in the same population at least for 3–10 million years^{22,23}. This time seems too long for neutral alleles¹⁵ and suggests that the polymorphism has been maintained largely by overdominant selection. It should be noted that heterozygote advantage may also arise from Clarke and Kirby's model of maternal-fetal incompatibility selection^{24,25}, although this model has been controversial²⁶.

Before the function of MHC molecules was clarified at the molecular level, several authors^{8,9} speculated that MHC alleles generate heterozygote disadvantage in association with infectious diseases and that in order to maintain a high level of polymorphism some kind of frequency-dependent selection is necessary. A popular model of frequency-dependent selection assumes that host individuals carrying a recently arisen mutant allele have a selective advantage because pathogens will not have had the time to adapt to infecting cells carrying a new mutant antigen^{8,9}. This model will result in a constant turnover of alleles in the population because old alleles lose resistance to pathogens. While this model is expected to generate a higher value of d_N than d_s , it can explain neither the high degree of

Table 2 Mean numbers of nucleotide substitutions per 100 synonymous sites (d_s) and per 100 nonsynonymous sites (d_N)

Human locus (No. comparisons)	Antigen recognition site (ARS) ($N = 57$)		Remaining codons in exons 2 and 3 ($N = 124$)		Exon 4 ($N = 92$)	
	d_s	d_N	d_s	d_N	d_s	d_N
<i>A</i> (45)	32.8 ± 8.8	39.3 ± 5.0	30.4 ± 6.0	12.9 ± 2.9**	28.7 ± 6.9	14.6 ± 2.5
<i>B</i> (36)	29.4 ± 8.0	36.6 ± 4.4	32.0 ± 6.0	14.4 ± 2.0**	33.9 ± 8.2	14.5 ± 2.6*
<i>C</i> (27)	27.7 ± 7.8	35.4 ± 4.8	36.5 ± 7.8	14.8 ± 2.0**	33.2 ± 8.2	14.9 ± 2.6*
Mean (108)	30.4 ± 8.0	37.4 ± 4.2	32.5 ± 5.8	13.9 ± 1.8**	31.6 ± 6.4	14.6 ± 2.5*
$d_s > d_N: d_N > d_s$	22:86		108:0		108:0	

N = Number of codons compared. The difference between d_s and d_N is significant at 5% level (*) and at 1% level (**).

polymorphism nor the persistence of polymorphic alleles at MHC loci mentioned above; a model of sequentially advantageous mutations similar to this is known to reduce the level of genetic polymorphism below that for neutral mutations²⁷. There are some other models of frequency-dependent selection²⁸, but they are not designed to explain all aspects of MHC polymorphism quantitatively. For these reasons, frequency-dependent selection seems to be less satisfactory than overdominant selection.

Although the above argument supports the overdominance hypothesis, the extent of heterozygote advantage (s) due to a single amino acid difference need not be very large. As long as the product of s and the long-term effective population size (N) is appreciably larger than 1, the extent of polymorphism is enhanced substantially¹². Since N is probably about 10^4 – 10^5 in humans and mice²⁹, s could be very small. This explains why some highly inbred wild populations can survive without much difficulty in the absence of MHC polymorphism^{30,31}.

The present study shows that comparison of synonymous and nonsynonymous substitutions is a useful way of studying positive darwinian selection. Hill and Hastie³² reported evidence supporting adaptive nucleotide substitution in functionally important regions of serine protease inhibitor genes. Our results are different as the selection detected is not directional but overdominant. If our conclusion is correct, it will be the first case of overdominant selection established at the molecular level since the discovery of the sickle-cell anaemia gene.

We thank Tommy Douglas, Manolo Gouy and Paul Sharp for their comments. This study was supported by grants from the NIH and the NSF.

Received 15 February; accepted 7 July 1980

1. Klein, J. *Natural History of the Major Histocompatibility Complex* (Wiley, New York, 1986).
2. Klein, J. *Adv. Immun.* **26**, 55–146 (1978).
3. López de Castro, J. A., Strominger, J. L., Strong, D. M. & Orr, H. T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3813–3817 (1982).
4. Ohta, T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3251–3254 (1982).
5. Mellor, A. L., Weiss, E. H., Ramachandran, K. & Flavell, R. A. *Nature* **306**, 792–795 (1983).
6. Doherty, P. C. & Zinkernagel, R. M. *Nature* **256**, 50–52 (1975).
7. Klein, J. & Figueroa, F. *CRC Crit. Rev. Immunol.* **6**, 295–386 (1986).
8. Snell, G. D. *Folia Biol.* **14**, 335–358 (1968).
9. Bodmer, W. F. *Nature* **237**, 139–145 (1972).
10. Björkman, P. J. *et al. Nature* **329**, 506–512 (1987).
11. Björkman, P. J. *et al. Nature* **329**, 512–518 (1987).
12. Maruyama, T. & Nei, M. *Genetics* **98**, 441–459 (1981).
13. Mellor, A. L. *Oxford Surveys Eukar. Gen.* **3**, 95–140 (1987).
14. Nei, M. & Gojobori, T. *Molec. Biol. Evol.* **3**, 418–426 (1986).
15. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
16. Holmes, N. & Parham, P. *EMBO J.* **4**, 2849–2954 (1985).
17. Hayashida, H. & Miyata, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2671–2675 (1983).
18. Tewarson, S., Zaleska-Rutczynska, Z., Figueroa, F. & Klein, J. *Tissue Antigens* **22**, 204–212 (1983).
19. Enssle, K.-H., Wagner, H. & Fleischer, B. *Human Immun.* **18**, 135–149 (1987).
20. Hedrick, P. W. & Thomson, G. *Genetics* **104**, 449–456 (1983).
21. Klitz, W., Thomson, G. & Bauer, M. P. *Am. J. hum. Genet.* **39**, 340–349 (1986).
22. McConnell, T. L., Talbot, W. S., McIndoe, R. A. & Wakeland, G. K. *Nature* **332**, 651–654 (1988).
23. Figueroa, F. & Klein, J. in *H-2 Genes, Products, and Function* (ed. David, C. S.) 61–76 (Plenum, New York, 1988).
24. Clarke, B. & Kirby, P. R. S. *Nature* **211**, 999–1000 (1966).
25. Hedrick, P. W. & Thomson, G. *Genetics* **119**, 204–212 (1988).
26. Gill, T. J. *Transplantation* **35**, 1–6 (1983).
27. Nei, M. in *Vito Volterra Symposium on Mathematical Models in Biology* (ed. Barigozzi, C.) 17–47 (Springer, Berlin, 1980).
28. Clarke, B. in *Genetic Aspects of Host-Parasite Relationships* (eds Taylor, A. E. R. & Muller, R.) 87–103 (Blackwell, Oxford, 1976).
29. Nei, M. & Graur, D. *Evol. Biol.* **17**, 73–118 (1984).
30. Figueroa, F., Tichy, H., Berry, R. J. & Klein, J. *Curr. Top. Microbiol. Immun.* **127**, 100–105 (1986).
31. O'Brien, S. J. *et al. Science* **227**, 1428–1434 (1985).
32. Hill, R. E. & Hastie, N. D. *Nature* **326**, 96–99 (1987).
33. Nei, M., Stephens, J. C. & Saitou, N. *Molec. Biol. Evol.* **2**, 66–85 (1985).
34. Koller, B. H. & Orr, H. T. *J. Immun.* **134**, 2727–2733 (1985).
35. Strachan, T., Sodoyer, R., Damotte, M. & Jordan, B. R. *EMBO J.* **3**, 887–894 (1984).
36. Cowan, E. P., Jelachich, M. L., Biddison, W. E. & Coligan, J. E. *Immunogenetics* **25**, 241–250 (1987).
37. N'Guyen, C., Sodoyer, R., Trucy, J., Strachan, T. & Jordan, B. R. *Immunogenetics* **21**, 479–489 (1985).
38. Zemmour, J., Ennis, P. D., Parham, P. & Dupont, B. *Immunogenetics* **27**, 281–287 (1988).
39. Szöts, H., Riethmüller, G., Weiss, E. & Meo, T. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1428–1432 (1986).
40. Kottman, A. H., Seeman, G. H. A., Guessow, H. D. & Roos, M. H. *Immunogenetics* **23**, 396–400 (1986).
41. Ways, J. P., Choppin, H. L. & Parham, P. *J. biol. Chem.* **260**, 11294–11933 (1985).
42. Gussow, D. *et al. Immunogenetics* **25**, 313–322 (1987).

43. Sodoyer, R. *et al. EMBO J.* **3**, 879–885 (1984).
44. Kvist, S., Roberts, L. & Dobberstein, B. *EMBO J.* **2**, 245–254 (1983).
45. Arnold, B., Burgert, H. C., Archibald, A. L. & Kvist, S. *Nucleic Acids Res.* **12**, 9473–9487 (1984).
46. Weiss, E. *et al. EMBO J.* **2**, 453–462 (1983).
47. Morita, T., Delarbre, C., Kress, M., Kourlisky, P., Gachelin, G. *Immunogenetics* **21**, 367–383 (1985).
48. Moore, K. W., Taylor, S. B., Sun, H., Eakle, K. A. & Hood, L. *Science* **215**, 679–682 (1981).
49. Schepart, B. S. *et al. J. Immun.* **136**, 3489–3495 (1986).
50. Watts, S. J. *Immun.* **139**, 3878–3885 (1987).
51. Taylor, S. B., Nairn, R., Coligan, J. E. & Hood, L. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1175–1179 (1985).
52. Duran, L. W., *et al. J. Immun.* **139**, 2818–2824 (1987).

Deletion of the human T-cell receptor δ -gene by a site-specific recombination

Jean-Pierre de Villartay*, Richard D. Hockett†, David Coran*, Stanley J. Korsmeyer & David I. Cohen*

* Laboratory of Chemical Biology, National Institute of Diabetes, and Digestive, and Kidney Diseases, National Institutes of Health, 9000 Rockville Pike, Bethesda, Maryland 20892, USA

† Departments of Medicine, Microbiology and Immunology, Howard Hughes Medical Institute, Washington University School of Medicine, St Louis, Missouri 63110, USA

The newly described T-cell receptor (TCR) δ locus^{1–5} is located inside the TCR α locus, between variable region (V) α and joining region (J) α ^{6,7}. Although the δ and α TCR genes are physically linked on the same chromosome, they are sequentially expressed during T-cell development⁸. This implies the existence of a highly efficient regulatory mechanism by which these two genes are independently rearranged. We have recently described⁹ a genetic element 'T early α ' (TEA) in humans transcribed in foetal thymocytes, spliced alternatively to constant region (C) α , and located between the TCR- δ locus (5') and the group of J α segments (3'). Importantly, TEA flanks a common site of rearrangement in the thymus, and distinguishes cells using TCR- γ/δ (TEA in germline configuration) from cells using TCR- α/β (TEA deleted on both chromosomes). In order to understand this TEA-associated recombination we analysed genomic clones representing these thymic rearrangements. We show that the TEA-associated recombination deletes the δ locus before productive (V δ D δ J δ) rearrangement. The diversity (D) δ and J δ regions, which provide the major source of δ gene diversity, are eliminated as a consequence of δ gene deletion and cannot then be used in conjunction with an α -TCR. We propose that the TEA-associated deletion of TCR- δ precedes the formation of an α -TCR and could down-regulate TCR- δ formation in maturing thymus.

A size-selected genomic thymus (GTH) library from DNA enriched for the predominant 4.5 kb SacI thymic rearrangement detected by the TEA element⁹ has been constructed in the SacI cloning site of the vector λ Zap (Stratagene). A complementary DNA (cDNA) containing a TEA element (Fig. 1a) was used to screen 500,000 recombinants, and eleven TEA-containing clones were isolated and classified into three groups according to their restriction enzyme pattern. Two groups contain only one member, GTH308 and GTH311, whereas the predominant group, called GTH 310, contains nine members (Fig. 1b). All three isolates contain a head-to-head join of two recombinational consensus sequences (heptamer-spacer-nanomer) as shown in Fig. 1c; consequently these three GTH isolates represent the excision product (reciprocal joint) of the TEA-associated rearrangement, rather than the direct rearrangement itself. These three clones are probably extrachromosomal circular DNA excised and possibly amplified following recombination as has been described for thymic rearrangements of the α - and β -TCR genes^{10,11}. All three clones contain identical sequences at one end but all diverge, after the heptamer -TCCTGTG- donated

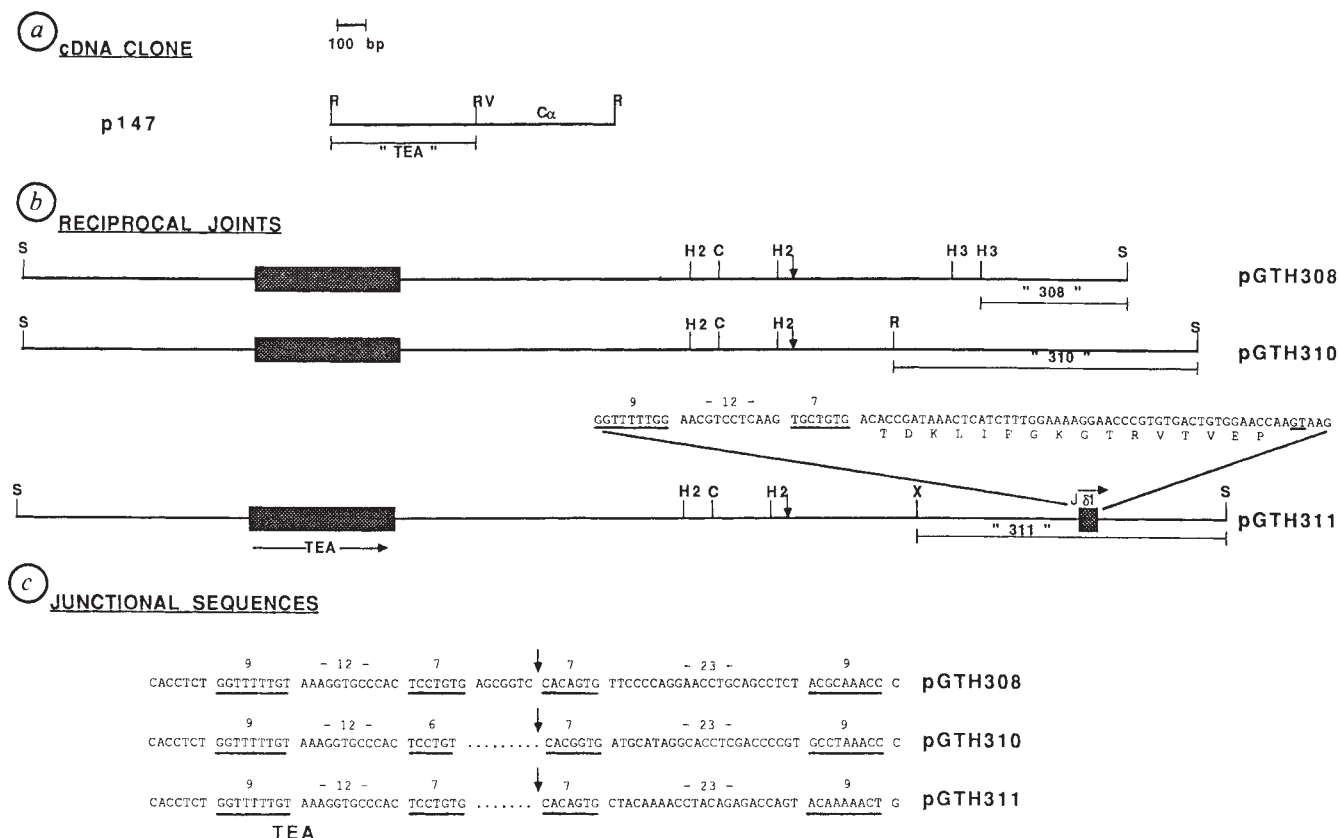


Fig. 1 Partial restriction map, sequence analysis and derivation of unique probes of TEA-associated reciprocal rearrangements. *a*, The *EcoRI-EcoRV* TEA fragment from the cDNA clone 147 (ref. 9) was used as a probe to isolate the subsequent thymic rearranged genomic clones. *b*, pGTH308, pGTH310 and pGTH311 represent the three different rearranged genomic clones isolated from the thymic size-selected genomic library corresponding to the 4.5 kb *SacI* band from Southern analysis¹⁸ of thymic DNA using the TEA probe. The inverted orientation placing TEA 5' to J δ in GTH311 indicates that these clones represent reciprocal joints and not direct rearrangement. This is also indicated by the nucleotide sequence (*c*) surrounding the recombination point (vertical arrow) which in all three cases consists of the head-to-head joining of the acceptor and donor recombination sequences whose heptamer and nanomer are underlined. The nucleotide sequence of J δ with its 5' heptamer-nanomer (underlined) reveals that this element is in the germline configuration in clone GTH311. This clone provided probes for J δ , allowing us subsequently to isolate full-length δ cDNAs including probes for C δ used in this report. The seven nucleotides between the two heptamers in GTH308 cannot be accounted for by the germline sequences (see Fig. 2*b*) and are probably due to N region addition as has been described in reciprocal recombinations of β -TCRs¹¹, GTH310 has an imprecise joining lacking one nucleotide from the acceptor heptamer. '308' (0.5 kb *SacI-HindIII*), '310' (1 kb *SacI-EcoRI*) and '311'/J δ (0.9 kb *SacI-XbaI*) represent the specific probes designed for further Southern analysis and the isolation of the germline counterpart. R = *EcoRI*, RV = *EcoRV*, S = *SacI*, H2 = *HindII*, C = *Clal*, H3 = *HindIII*, X = *XbaI*.

Methods. For construction of the size-selected library, 100 μ g of high molecular weight, 3-year-old human thymic DNA was digested to completion with a 4-fold excess of *SacI* restriction endonuclease and run in a 0.7% Seakem agarose gel. A slice was cut out of the gel, corresponding to 4.5 kb and the DNA eluted before cloning in the *SacI* site of the vector λ Zap (Stratagene) using standard techniques¹⁹. The TEA probe was used to screen 5×10^5 recombinant phage, from which 11 clones were isolated and grouped in three categories according to restriction maps before subcloning in pGEM4 (Promega). Sequencing was by the dideoxy termination method of Sanger²⁰ on double-stranded DNA²¹ using SP6 and T7 specific oligonucleotides. Sequences were extended by the generation of 3' internal oligonucleotides.

by the TEA region (Fig. 1*c*) so these three clones recombine at the same acceptor site, but at three different donor sites. We located the unrearranged J δ segment within GTH311 (Fig. 1*b*), showing that this excision product eliminated essential parts of the δ locus when it is not used. As reciprocal recombinants do not retain the donor and acceptor elements which are joined in the direct recombination, we obtained these regions from germline (GL) sequences. Clones GL TEA (λ 111A/p16), GL308 (pFB84.1), GL310 (p501.12) and GL311 (p311HR3) represent the germline counterparts of TEA, GTH308, GTH310 and GTH311, respectively (Fig. 2*a,b*). The sequence of these germline clones could then be compared with their rearranged counterparts, and the point at which the two sequences diverge is then defined as the donor or acceptor site.

The sequence of germline TEA (GL TEA) just adjacent to the recombination site (GL TEA/p16 in Fig. 2*b*) reveals that the 3' acceptor site for rearrangement is highly homologous to a *J α* segment, with its heptamer-nanomer sequence on the

5' side and an open reading frame and a donor splice sequence on its 3' side. It is likely to represent a *J α* pseudogene because, despite the presence of several conserved amino acids, the *J α* consensus sequence FGXGT is not complete in GL TEA (N in place of the second G), which lacks in addition three other residues (L, V, P) conserved in at least 70% of *J α s*. In addition a homology search between GL TEA and 44 *J α s* from functional cDNA clones¹²⁻¹⁴ shows that the GL TEA element has not been reported, also implying that it is a pseudogene. The germline genomic phage spanning this region (λ 111A in Fig. 2*a*) contains *J α* segments; moreover, the acceptor site of TEA-associated recombination is likely to be in front of any *J α* segment which is used, because we could find no additional *J α s* in the 2.5 kb of sequence immediately upstream. The TEA-flanking acceptor site representing the 3' recombining element is called $\Psi J $\alpha$$, and a probe for this region is described in Fig. 2*a*.

We isolated the germline form of the principal 5' site (donor site) of rearrangement (GL310, Fig. 2*b*) and defined the donor,

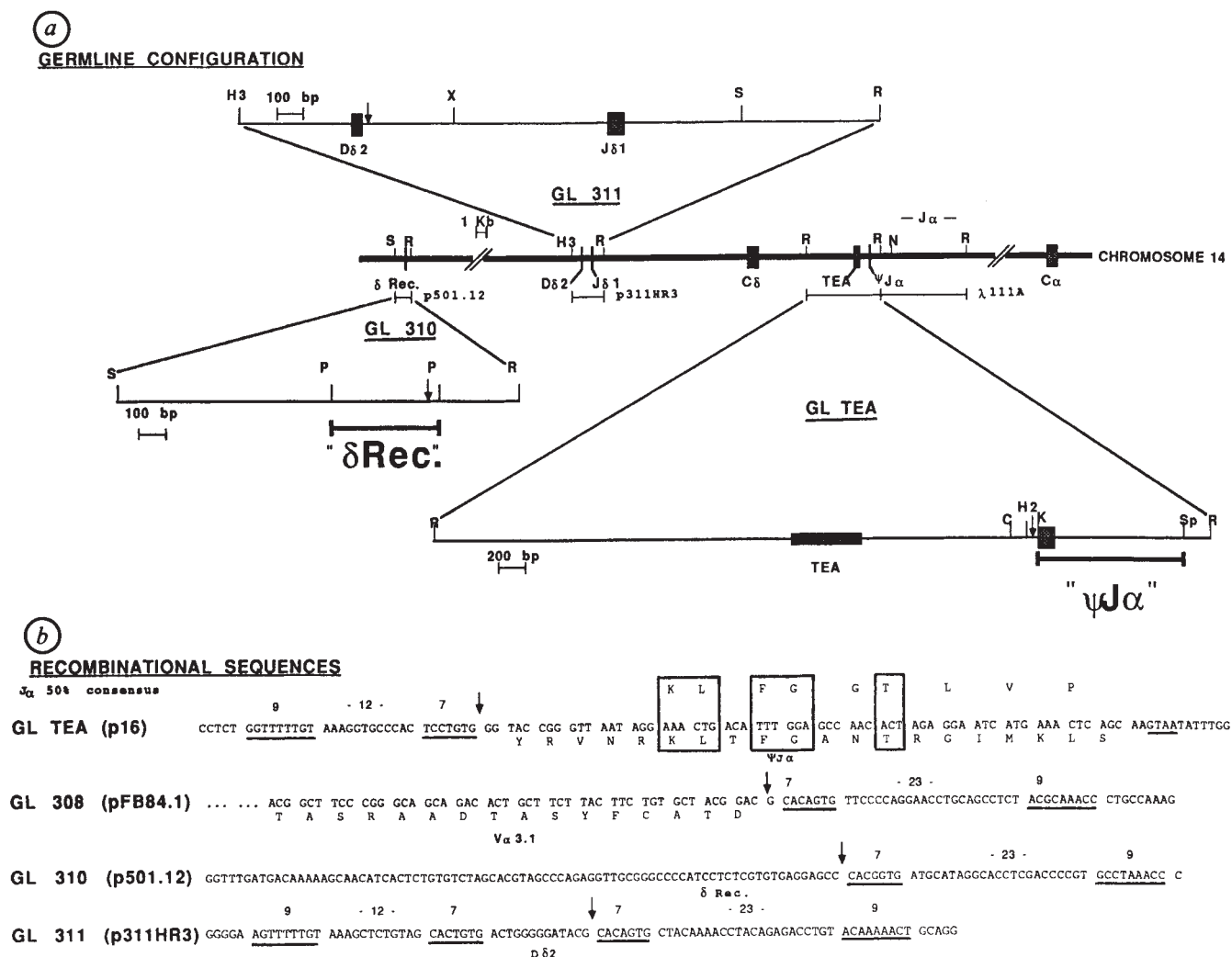


Fig. 2 Partial restriction map, sequence analysis and derivation of unique probes corresponding to the germline (GL) donor and acceptor sites of TEA-associated direct recombination. The overall map generated with these clones is shown in bold face. *a*, GL clones were isolated as λ phage recombinants using probes described in Fig. 1*a,b*. GL TEA (λ 111A) is from an *Mbo*I human placental library in Charon 28, the generous gift of Dr Philip Leder; the other isolates are from a human lung fibroblast library prepared in the vector λ Fix (Stratagene). p16, pFB84.1, p501.12 and p311HR3 represent subclones in the pGEM4 plasmid derived from the germline clones GL TEA, GL308, GL310 and GL311 respectively and used in sequence analysis. In GL311 the precise orientation and position of D δ 2 and J δ has been determined by sequencing the region spanning these two elements. The map of this region of DNA was determined by genomic walking, PFGE and deletional analysis (see text). *b*, For clarity only the sequence surrounding the recombination site (vertical arrow) is indicated. Heptamer and nanomer recombination sequences are underlined. The sequence homology between ψ J α and the 50% consensus J α sequence is shown in open boxes. ' ψ J α ' (1 kb *Sph*I-*Kpn*I) and ' δ Rec' (0.4 kb *Pvu*II-*Pvu*II) probes have been generated for further Southern blot analysis. R = *Eco*RI, S = *Sac*I, H2 = *Hind*II, C = *Cla*I, H3 = *Hind*III, X = *Xba*I, Sp = *Sph*I, N = *Nco*I, P = *Pvu*II and K = *Kpn*I.

termed δ Rec, as a sequence containing at its 3' end a consensus heptamer-nanomer with the appropriate 23 base pair (bp) spacer to permit recombination to ψ J α (Fig. 2*b*). Since δ Rec lacks a meaningful open reading frame, its rearrangement to ψ J α cannot encode a functional TCR protein. Two additional donor sites representing the infrequent reciprocal recombinations, GTH311 and GTH308 (Fig. 1*b*), were also isolated (GL311, GL308, Fig. 2*b*). The GL311 donor is defined as human D δ 2 (Fig. 2*b*), being identical at 10 of 13 bases with the sequence of murine D δ 2⁴. The third donor (GL308, Fig. 2*b*) is a V α gene (V α 3.1)¹². The GL311 recombination leads to the deletion of J δ and C δ which lie between D δ 2 and ψ J α . This deletion is documented by the presence of the unrearranged J δ segment in the excisional product GTH311. Also, as all productive δ -TCR rearrangements in human and mouse appear to use D δ 2¹⁻⁵, '311-recombination' is a mutually exclusive alternative to forming an intact δ -TCR gene.

When we used ψ J α and δ Rec as probes to analyse human DNA (Fig. 3), we found that the germline bands (represented

by the B cell line SG1) for ψ J α (Fig. 3*a*) and δ Rec (Fig. 3*b*) are different in size. However, both probes detect an identical 4.8 kb *Xba*I rearrangement occurring frequently and specific to thymic DNA (Fig. 3*a, b*, thymus), reflecting the major thymic rearrangement that aligns δ Rec and ψ J α . Since splenic DNA does not retain this rearrangement (Fig. 3), it must be an intermediate event specific for developing thymocytes. Only one major rearranged band is detected, confirming previous results obtained with the TEA probe⁹. To study the consequences of δ Rec/ ψ J α rearrangement, we analysed the DNA from twenty-five acute leukaemic T cells (T-ALL) and lines. One T-ALL (376)¹⁵ and one cell line (HSB)¹⁶ show particularly informative rearrangements with both the δ Rec and ψ J α probes. A CD3, CD4 and CD8-bearing cell, 376, expresses the α/β receptor on its surface. The HSB cell line does not carry these antigens and since it does not express CD3, it does not produce a TCR. In both cells the germline 5.8 kb *Xba*I band corresponding to δ Rec and the 6.2 kb *Xba*I band corresponding to ψ J α are deleted (Fig. 4*b,e*). Both HSB and 376 show the same rearranged 4.8 kb

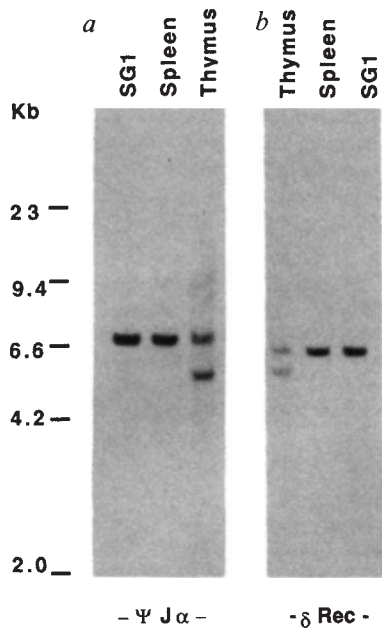


Fig. 3 Southern blot analysis of the direct rearrangements associated with TEA. 10 µg of human DNA from adult thymus (3 yr old), spleen and an Epstein-Barr virus-transformed B cell line (SG1) were digested with *Xba*I, electrophoresed on a 0.7% agarose gel and blotted on to nylon membranes (Nytran, Schleicher & Schuell). The *a* and *b* lanes were run in parallel on the same gel. Hybridization was performed under standard conditions (40% formamide/5×SSPE, 42°C) followed by moderate stringency washes (0.2×SSPE/0.1% SDS, 55°C). (1×SSPE is 0.18 M NaCl, 0.01 M sodium phosphate, pH 7.4 and 1 mM EDTA)¹⁹. *ψJα* and *δRec* probes, depicted in Fig. 2a, were radiolabelled by random priming²². The rearranged band revealed by the *ψJα* probe, *a*, and the *δRec* probe, *b*, are identical in size (4.8 kb). Human thymus aged 7 months, 10 months and 2 yr reveals the same rearrangement. Size markers (kb) show positions of *Hind*III fragments of λ DNA.

*Xba*I band, previously observed in thymic DNA (Fig. 3) with either *δRec* or *ψJα*. The absence of *Dδ*, *Jδ* and *Cδ* hybridization to the DNA of either cell (Fig. 4c, d, data not shown) leads to the conclusion that each has deleted the *δ* gene on both chromosomes. Both of these cells and several *α/β* cells retain *Vδ1* in the germline configuration (Fig. 4a), indicating that the principal *Vδ* gene was not rearranged before the deletion. Furthermore, of nine *γ/δ*-expressing T cells, none showed *δRec* rearrangement (data not shown) which distinguishes this event from those occurring in cells expressing *γ/δ*-TCRs. Altogether these results establish that *δRec/ψJα* recombination leads to the deletion of *Dδ*, *Jδ* and *Cδ*; *δRec* being the 5' and *ψJα* the 3' deleting elements for TCR-*δ*. Figure 2a shows the relative position of *δRec* and the precise position of *ψJα*. Pulsed field gel electrophoresis (PFGE) maps *δRec* to the same 180 kb *Sfi*I fragment as *Dδ*, *Jδ* and *ψJα*, but to a 40 kb *Xho*I fragment just 5' to these elements (data not shown) so that its 5' position leads to *Dδ*, *Jδ*, and *Cδ* deletion upon rearrangement to *ψJα*¹⁷.

These data establish that a common thymic rearrangement joining a 5' deleting element (*δRec*) to the 5' end of the *Jas* (*ψJα*) leads to the deletion of TCR *δ*. Cells undergoing *δRec* rearrangement on both chromosomes cannot bear *γ/δ*-TCRs, and these cells could represent a terminal pathway or an intentional intermediate step in functional differentiation. If the deletion rearrangement is a terminal step, the cell dies without ever producing a functional *α/β*- or *γ/δ*-TCR and therefore could not undergo thymic selection ('self-education'). It seems very unlikely that such a common deletion mechanism could exist without any apparent biological function. Instead, we propose a model of T-cell ontogeny in which productive *δ*-TCR rearrangement on the one hand and *δ* gene deletion on the other would be two events separating distinct lineages of T cells bearing either *γ/δ*- or *α/β*-TCRs. In the *γ/δ*-TCR lineage, the initial recombination of *Vδ* to *Dδ* (ref. 4) is an unusual pattern of assembly among immunoglobulin or TCR genes and would lead to the deletion of *δRec*, at least on one chromosome. In contrast, in the *α/β* TCR lineage, as a consequence of *δRec/ψJα* recombination, the *Dδ*, *Jδ*, and *Cδ* genes are

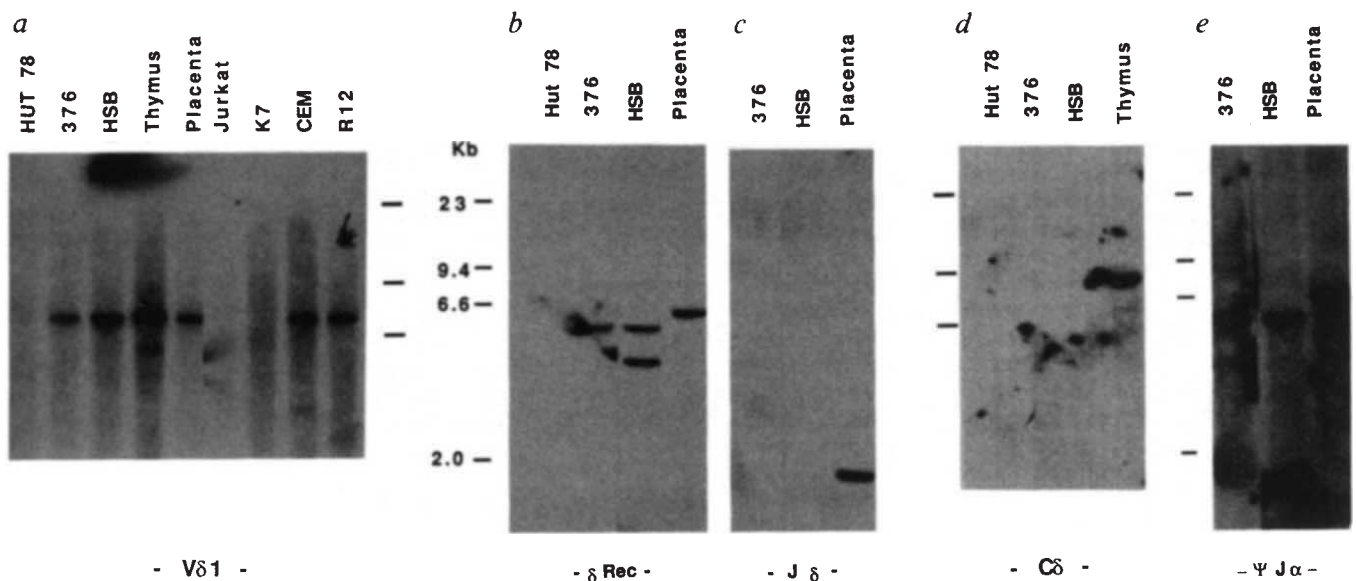


Fig. 4 Localization of *δRec* by Southern deletional analysis. Experimental conditions were identical to those described in Fig. 3. Hut 78, Jurkat²³, K7²⁴, CEM and R12²⁴ are *α/β*-TCR-bearing mature human T cell lines. DNA was digested with *Xba*I restriction endonuclease. *Jδ* probe (c) is depicted in Fig. 1b. *δRec* (b) and *ψJα* (e) probes are presented in Fig. 2a. *Vδ1* probe (a) is a 1.0 kb *Pst*I fragment isolated from a genomic DNA clone (not shown), representing the widely used human *Vδ* sequence described to date as part of *δ* cDNA^{1,2} and detects an 8.5 kb germline band. *Cδ* probe (d) is a 1.4 kb *Eco*RI fragment isolated from a TCR-*δ* cDNA clone (not shown). Size markers are identical to those in Fig. 3. In the case of HSB, the second rearranged band detected by the *δRec* probe (b) probably represents rearrangement to a downstream *Jα* segment which would have the effect of removing *ψJα* on this chromosome. In reference 15, fresh T-ALL 376 is referred to as case 6.

removed, and the principal human δ diversifying elements, D δ and J δ , could not be used in conjunction with the α -TCR. This could lead to α/β - and γ/δ -TCRs having different recognition requirements. Following δ gene deletion, further V α to J α rearrangements would be needed to construct a functional α -TCR gene. δ Rec/ΨJ α rearrangement clearly precludes TCR- δ gene usage and at this point we favour the model that it defines an intermediate event in the process of differentiation toward the α/β lineage.

We acknowledge Drs William Paul and Warren Leonard for critical review of this manuscript, and Drs Erwin Tschachler, Michael Davey and Jeffrey Cossman for materials used in this report.

Received 5 April; accepted 6 July 1988.

- Hata, S., Brenner, M. B. & Krangel, M. S. *Science* **238**, 678-681 (1987).
- Loh, E. Y. *et al. Nature* **330**, 569-572 (1987).
- Chien, Y., Iwashima, M., Kaplan, K. B., Elliott, J. F. & Davis, M. M. *Nature* **327**, 677-682 (1987).
- Chien, Y. *et al. Nature* **330**, 722-727 (1987).
- Elliott, J. F., Rock, E. P., Patten, P. A., Davis, M. M. & Chien, Y. *Nature* **311**, 627-631 (1988).
- Boehm, T. *et al. EMBO J.* **7**, 385-394 (1988).
- Croce, C. M. *et al. Science* **227**, 1044-1047 (1985).
- Pardoll, D. M. *et al. Nature* **326**, 79-81 (1987).
- de Villartay, J. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 8608-8612 (1987).
- Fujimoto, S. & Yamagishi, H. *Nature* **327**, 242-243 (1987).
- Okazaki, K., Davis, D. D. & Sakano, H. *Cell* **49**, 477-485 (1987).
- Yoshikai, Y., Kimura, N., Toyonaga, B. & Mak, T. W. *J. exp. Med.* **164**, 90-103 (1986).
- Klein, M. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 6884-6888 (1987).
- Kimura, N., Toyonaga, B., Yoshikai, Y., Du, R. P. & Mak, T. W. *Eur. J. Immunol.* **17**, 375-383 (1987).
- Pittaluga, S., Raffeld, M., Lipford, E. H. & Cossman, J. *Blood* **68**, 134-139 (1986).
- Minowada, J. in *Leukemia* (eds Gunz, F. W. & Henderson, E. S.) 119-139 (Grune & Stratton, New York, 1983).
- Siminovich, K. A., Bakshi, A., Goldman, R. & Korsmeyer, S. J. *Nature* **316**, 260-262 (1985).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1983).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Zagursky, R. J., Baumeister, K., Lomax, N. & Berman, M. *Gene Anal.* **2**, 89-94 (1985).
- Feinberg, A. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).
- Weiss, A. & Stobo, J. *J. exp. Med.* **160**, 1284-1299 (1984).
- Mitsuya, H. *et al. Science* **223**, 1293-1296 (1984).

Intrathymic deletion of self-reactive cells prevented by neonatal anti-CD4 antibody treatment

H. Robson MacDonald*, Hans Hengartner† & Thierry Pedrazzini‡

* Ludwig Institute for Cancer Research, Lausanne Branch, 1066 Epalinges, Switzerland

† Institute of Pathology, University Hospital, 8091 Zurich, Switzerland

‡ World Health Organization Immunology Research and Training Center, Institute of Biochemistry, 1066 Epalinges, Switzerland

T-cell differentiation in the thymus involves the coordinate expression of genes encoding the α and β chains of the major histocompatibility complex-restricted heterodimeric antigen receptor (TCR) complex, as well as other functionally important molecules such as CD4 and CD8. The repertoire of TCR expressed by T cells is generally thought to be influenced by positive and/or negative selection events occurring when TCRs on developing T cells interact with self-antigens and major histocompatibility complex components¹. Using a model system in which specific antigen-reactive cells can be monitored by virtue of their preferential expression of certain TCR β -chain variable (V β) domains, it has been shown²⁻⁴ that self-reactive T cells are clonally deleted during development. We report here that clonal deletion of V β_{66} cells in Mls^a mice can be prevented by *in vivo* neonatal administration of monoclonal antibodies directed against CD4. Furthermore, as

Table 1 Binding of anti-CD4 and anti-CD8 mAbs to thymocytes *in vivo*

Administered mAb <i>in vivo</i> (μ g per mouse)	Staining mAb	% Positive thymocytes	Mean fluorescence intensity (arbitrary units)
PBS (-)	GK-1.5	92	120
	None	3	—
GK-1.5 (1,000)	GK-1.5	86	106
	None	83	102
PBS(-)	H35-17.2	89	141
	None	0.2	—
H35-17.2 (100)	H35-17.2	83	121
	None	85	117

Newborn (4-day-old) Balb/c mice were injected intraperitoneally with phosphate-buffered saline or with the indicated doses of partially purified mAbs GK-1.5¹⁶ (anti-CD4) or H35-17.2¹⁷ (anti-CD8). After 4 h thymocytes were prepared and stained at 4 °C with fluoresceinated goat anti-rat immunoglobulin alone or with a saturating concentration of the relevant mAb (100 μ g ml⁻¹), followed by fluorescent anti-rat immunoglobulin. Data are presented both as % positive cells (relative to controls from untreated mice) and as mean fluorescence intensity (arbitrary units) of the positive cells. The slightly reduced staining of thymocytes from mAb-treated animals probably results from modulation of CD4 and CD8 molecules *in vivo*. Background staining was <10 arbitrary units.

anti-CD4 monoclonal antibody treatment resulted in the reappearance of V β_{66} cells in the mature CD8⁺ T-cell subset, it is likely that clonal deletion acts on the CD4⁺CD8⁺ thymocyte subset and that this subset is an intermediate stage in the differentiation pathway of both CD4⁺ and CD8⁺ T-cell lineages.

T-cell receptors which use the V β_6 (ref. 2) or V $\beta_{8.1}$ (ref. 4) gene segments react preferentially with antigens encoded by the Mls^a allele of the minor lymphocyte stimulatory (Mls) locus, and moreover V β_{66} or V $\beta_{8.1}$ (thought to be self-reactive) T cells are deleted in mouse strains that express Mls^a. An unexpected observation in our study was that CD8⁺ (as well as CD4⁺) T cells carrying V β_{66} were deleted, even though only the CD4⁺ subset could be shown to be reactive towards Mls^a. To explain this apparent anomaly, we postulated that CD8⁺V β_{66} T cells were derived from CD4⁺ precursors that had the capacity to recognize Mls^a (and hence be deleted during development) and that a candidate for this precursor cell was the CD4⁺CD8⁺ thymocyte. To test this hypothesis more directly, we have now treated neonatal Mls^a mice with monoclonal antibodies (mAbs) directed against the CD4 molecule in an effort to subvert the intrathymic deletion process.

Figure 1 shows the selective depletion of V β_{66} and V $\beta_{8.1}$ T cells in control congenic Balb/c mouse strains which differ only at the Mls^a locus⁵. Quantitation of V β_{66} cells was accomplished directly^{6,7}, whereas measurement of V $\beta_{8.1}$ and V $\beta_{8.2}$ cells was carried out by a double-staining protocol using anti-V $\beta_{8.1/8.2}$ (refs 8, 9) and anti-V $\beta_{8.2}$ (ref. 4) mAbs. The latter method allows direct analysis of V $\beta_{8.1}$ (as well as of V $\beta_{8.2}$) cells, avoiding more complicated subtractive procedures for V $\beta_{8.1}$ quantitation⁴. V β_{66} and V $\beta_{8.1}$ cells were severely depleted among mature cortisone-resistant thymocytes of congenic Mls^a mice, whereas V $\beta_{8.2}$ cells were present at levels comparable to control Balb/c mice. Cortisone-resistant thymocytes were chosen as indicator cells for these (and all subsequent) experiments because their T-cell population is more than 95% mature (CD4- or CD8-bearing)¹⁰ and can easily be obtained as early as 1 week after birth, shortening the *in vivo* treatment protocols.

The effect of neonatal treatment with anti-CD4 or anti-CD8 mAbs was then investigated. Preliminary studies (Table 1) established the mAb doses sufficient for saturation binding to total thymocytes (most of which are CD4⁺ CD8⁺) *in vivo*. When cortisone-resistant thymocytes from mAb-treated mice were examined (Table 2), the anti-CD8 treatment effectively depleted

CD8⁺ mature T cells, whereas the anti-CD4 treatment resulted in only a modest reduction in the corresponding CD4⁺ subset. Neither treatment demonstrably eliminated mature T cells of the other phenotype. It should be noted that these data (which will be discussed in detail elsewhere¹¹) conflict with those in a recent study¹², suggesting that CD4⁺ mature T cells are depleted by *in vivo* treatment with anti-CD8 mAbs in a radiation chimera model.

When Mls^a congenic Balb/c mice were treated from birth with anti-CD4 mAbs, a dramatic (about tenfold) and reproducible increase in both V_{β6}⁺ and V_{β8.1}⁺ cortisone-resistant thymocytes was observed, reaching levels of about 40% and 70% respectively of those observed in control Balb/c mice (Table 2). Mls^a mice treated with functionally equivalent amounts of isotype-matched rat mAb directed against CD8 or the hapten 2,4-dinitrophenyl (DNP) did not differ from untreated controls in their pattern of TCR expression, so the increase was related to the specific mAb treatment (Table 2). Furthermore, increased expression of V_{β6} and V_{β8.1} in the anti-CD4 mAb-treated mice was selective in that expression of V_{β8.2} (a gene segment not associated with Mls^a reactivity⁴) actually decreased slightly compared to controls (presumably because of random redistribution of TCR expression in the absence of clonal deletion). The re-appearance of V_{β6}⁺ and V_{β8.1}⁺ mature T cells in anti-CD4 (but not anti-CD8) mAb-treated Mls^a mice indicates that the CD4 molecule is of critical functional importance in the recognition of Mls^a antigen (and/or associated MHC class II gene products) during the process normally leading to clonal deletion of self-reactive cells. In this regard, the more extensive reconstitution (proportionately) of V_{β8.1}⁺ cells compared to V_{β6}⁺ cells following mAb treatment could be interpreted as suggesting that V_{β8.1}⁺ TCRs have lower intrinsic affinity for Mls^a/class II MHC determinants than do V_{β6}⁺ TCRs, thereby allowing V_{β8.1}⁺ cells to escape elimination more readily when CD4 function is impaired.

Clonal deletion of V_{β6}⁺ cells occurs in both CD4⁺ and CD8⁺ mature T cell subsets in Mls^a mice²: we therefore examined the phenotype of surviving V_{β6}⁺ cells in mAb-treated mice using double-staining procedures. (V_{β8.1} expression was not examined in this way because three-colour flow microfluorometry was not available). V_{β6}⁺ cells could be found in both CD8⁺ and CD8⁻ subsets of cortisone-resistant thymocytes from the anti-CD4 mAb treated Mls^a mice, whereas (as expected) no detectable V_{β6}⁺ cells were seen in either subset from control or anti-DNP mAb-treated animals (Fig. 2). The presence of CD8⁺ V_{β6}⁺ mature T cells in the anti-CD4 mAb-treated mice provides compelling evidence that these cells arise from a CD4⁺ precursor (although the formal possibility that CD4⁺ cells control the development of CD8⁺ cells indirectly cannot be excluded). We believe that this precursor is most likely to be a CD4⁺CD8⁺ thymocyte. According to this view, the 'rescue' of CD8⁺ V_{β6}⁺ T cells following *in vivo* anti-CD4 mAb treatment would result from interference

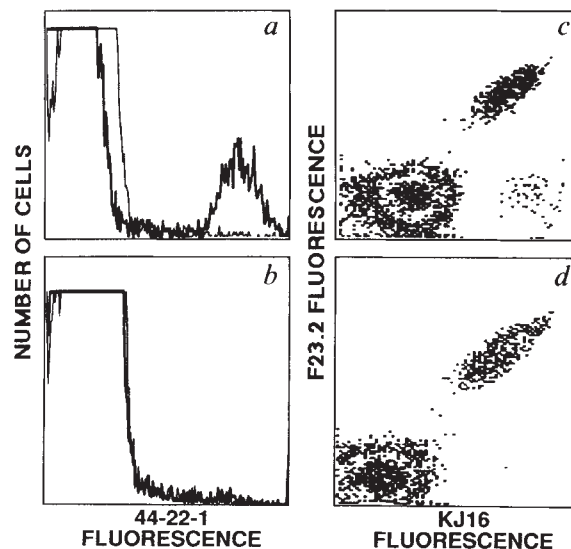


Fig. 1 Selectively reduced expression of V_{β6} and V_{β8.1} in mature T cells from Mls^a congenic Balb/c mice. Cortisone-resistant thymocytes from Balb/c (a) or congenic (Balb/c × Balb.D2.Mls^a)F₁ (b) mice were stained with anti-V_{β6} mAb 44-22-1 followed by fluoresceinated goat anti-rat immunoglobulin. Alternatively (c and d), cells were double-stained with anti-V_{β8.1/8.2} mAb KJ16 (revealed with fluorescent goat anti-rat immunoglobulin) and anti-V_{β8.2} mAb F23.2 (revealed with phycoerythrin-conjugated goat anti-mouse immunoglobulin G₁). The latter staining reveals V_{β8.1}⁺ and V_{β8.2}⁺ cells as discrete populations (lower right and upper right, respectively), as F23.2 does not block the subsequent binding of KJ16 (data not shown).

Methods. Mice (4–5 weeks old) were injected intraperitoneally with hydrocortisone acetate (4 mg per mouse) 48 h before they were killed. mAbs 44-22-1, KJ16 and F23.2 were used as undiluted hybridoma culture supernatants. Fluorescein isothiocyanate (FITC)-conjugated goat anti-rat immunoglobulin and phycoerythrin (PE)-conjugated goat anti-mouse immunoglobulin G₁ were obtained from TAGO (Burlingame, California) and Southern Biotech (Birmingham, Alabama) respectively. For staining, 5 × 10⁵ cortisone-resistant thymocytes in 100 μl were incubated sequentially for 30 min at 4 °C with first antibody, FITC conjugate, and (if necessary) second antibody and PE conjugate. All samples were passed on a modified FACS II flow cytometer gated to exclude non-viable cells¹⁹. 10⁴ viable cells were accumulated for histograms (2 × 10⁴ for cytograms) using logarithmic amplification of fluorescence intensity. Background staining (thin lines) is illustrated for the histograms.

Table 2 CD4/CD8 phenotype and use of TCR V_β in control and mAb-treated Mls^a congenic Balb/c mice

Strain	mAb treatment	% Positive cells				
		CD4	CD8	V _{β6}	V _{β8.1}	V _{β8.2}
Balb/c	None	71.4, 66.0	28.5, 23.0	10.2, 10.0	6.2, 7.0	14.0, 14.4
Balb.D2.Mls ^a F ₁	None	76.8, 64.2	23.2, 20.0	0.3, 0.4	0.5, 0.6	16.1, 15.3
Balb.D2.Mls ^a F ₁	Anti-DNP	79.5	20.6	0.2	0.5	15.0
Balb.D2.Mls ^a F ₁	Anti-CD4	58.6 ± 4.9	41.2 ± 4.6	3.9 ± 0.4	4.5 ± 0.3	10.7 ± 0.4
Balb.D2.Mls ^a F ₁	Anti-CD8	93.7 ± 0.3	2.7 ± 0.5	0.2 ± 0.3	1.2	12.3

(Balb/c × Balb.D2.Mls^a)F₁ mice were injected twice daily from birth with partially purified rat immunoglobulin G2b mAbs directed against CD4 (GK-1.5)¹⁶, CD8 (H35-17.2)¹⁷ or DNP (LO-DNP-57)¹⁸. Anti-CD4 and anti-DNP mAbs were used at 1 mg per injection; anti-CD8 mAbs were used at 100 μg per injection. On days 6 and 7, hydrocortisone acetate (4 mg per mouse) was injected. Cortisone-resistant thymocytes were recovered on day 8 and stained with mAbs GK-1.5 (anti-CD4), H35-17.2 (anti-CD8), 44-22-1 (anti-V_{β6}), KJ16 (anti-V_{β8.1/8.2}) and F23.2 (anti-V_{β8.2}). Data are presented as % positive cells (after subtraction of background staining with the appropriate fluorescent anti-immunoglobulin reagent). Independent experiments (separated by commas) represent pools of 2–5 mice, and standard deviations are given where 3–5 individual mice were analysed in a single experiment.

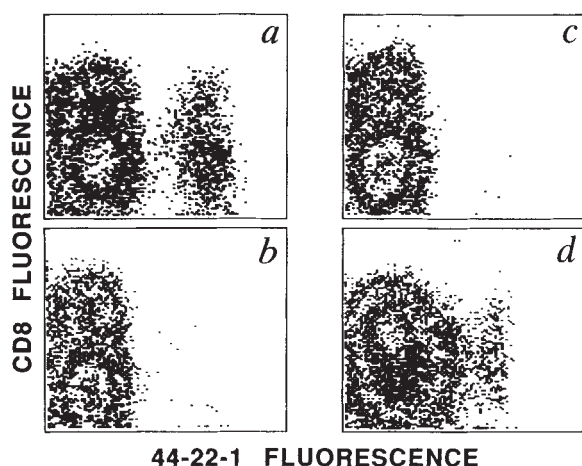


Fig. 2 Correlation of CD8 and $V_{\beta 6}$ expression on cortisone-resistant thymocytes from anti-CD4-treated and control Mls^a congenic Balb/c mice. (Balb/c $\times$ Balb.D2.Mls^a) F_1 mice were treated from birth with anti-CD4 (d) or control anti-DNP mAbs (c) and injected with hydrocortisone. Recovered cortisone-resistant thymocytes were double-stained with anti- $V_{\beta 6}$ mAb 44-22-1 and anti-CD8 mAb H35-17.2. Cortisone-resistant thymocytes from untreated Balb/c (a) and (Balb/c $\times$ Balb.D2.Mls^a) F_1 (b) mice are included for comparison. CD8⁺ cells accounted for 33.1%, 23.2%, 20.6% and 46.5% of the total in a–d respectively. The proportion of $V_{\beta 6}$ cells in the corresponding subsets was 10.6% (CD8⁺) and 10.7% (CD8[–]) in a, and 4.1% (CD8⁺) and 4.3% (CD8[–]) in d.

Methods. Neonatal injections and cortisone treatment were as in Table 1. Staining with mAb 44-22-1 (see Fig. 1) was followed by biotinylated anti-CD8 mAb H35-17.2 (2 μ g) and PE-conjugated avidin (10 μ g). Samples were analysed as in Fig. 1.

with the accessory function of CD4 (acting either as an adhesion-strengthening molecule¹³ or as part of the functional TCR complex¹⁴) during the interaction of TCRs on $V_{\beta 6}$ CD4⁺CD8⁺ precursors with Mls^a/class II MHC molecules on thymic non-lymphoid cells. In the most straightforward model, such interference would neutralize the negative selection mechanism and thereby allow further differentiation into $V_{\beta 6}$ CD8⁺ mature T cells. Alternatively, because Mls^a reactivity is strongly correlated with CD4 expression¹⁵, it is possible that a positive selection process in Mls^a mice diverts most $V_{\beta 6}$ CD4⁺CD8⁺ precursors to the CD4 lineage before negative selection. In this situation, anti-CD4 mAb treatment would block the dominant positive selection and again allow $V_{\beta 6}$ CD8⁺ mature T cells to develop.

The nature of the CD8⁺ $V_{\beta 6}$ cells 'rescued' by neonatal anti-CD4 mAb treatment requires further study. Although these cells are likely to be CD4⁺ (CD4⁺ and CD8⁺ cells account for more than 95% of cortisone-resistant thymocytes in normal mice¹⁰), the possibility that they belong to an aberrant (cortisone-resistant) subset of CD4[–] thymocytes cannot be excluded. If indeed CD4⁺ $V_{\beta 6}$ cells do arise after anti-CD4 mAb treatment, then clonal deletion can be subverted without concomitant loss of the CD4⁺ T-cell lineage (because putative positive selection mechanisms requiring CD4 are blocked). The mechanism underlying this puzzling dichotomy is unclear, but could be related to differing affinity thresholds for interactions between TCRs and/or CD4 with their corresponding ligands during negative and presumed positive selection events.

We demonstrate that clonal deletion of self-reactive cells (at least in the Mls^a system) can be prevented by neonatal administration of anti-CD4 mAbs. Furthermore, subject to the caveats mentioned above, our results indicate that the primary target of anti-CD4 mAbs in this system is the CD4⁺CD8⁺ thymocyte subset. These results are therefore compatible with the idea that physiological T-cell tolerance occurs at the CD4⁺CD8⁺ stage of development and that CD4⁺CD8⁺ thy-

mocytes contain the precursors of both the CD4⁺ and CD8⁺ mature T-cell lineages.

We wish to thank Andrée Porret and Nathalie Jeanguenat for technical assistance, Drs J. Kappler, U. Staerz, F. Fitch, M. Pierres and H. Bazin for the mAbs and Pierre Zaech for performing the flow cytometry. This work was supported in part by a grant (to H.H.) from the Swiss National Science Foundation.

Received 13 June; accepted 11 July 1988.

1. Moller, G. *Immunol. Rev.* **101**, 1–215 (1988).
2. MacDonald, H. R. *et al. Nature* **332**, 40–45 (1988).
3. Kappler, J. W., Roehm, N. & Marrack, P. *Cell* **49**, 273–280 (1987).
4. Kappler, J. W., Staerz, U., White, J. & Marrack, P. *Nature* **332**, 35–40 (1988).
5. Festenstein, H. & Berumen, L. *Transplantation* **37**, 322–324 (1984).
6. Acha-Orbea, H., Zinkernagel, R. M. & Hengartner, H. *Eur. J. Immun.* **15**, 31–36 (1985).
7. Payne, J. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
8. Haskins, K. *et al. J. exp. Med.* **160**, 452–471 (1984).
9. Behlke, M. A. *et al. J. exp. Med.* **165**, 257–262 (1987).
10. Ceredig, R., Dialynas, D. P., Fitch, F. W. & MacDonald, H. R. *J. exp. Med.* **158**, 1654–1671 (1983).
11. MacDonald, H. R. *et al. Immunol. Rev.* **104**, 157–182 (1988).
12. Smith, L. *Nature* **326**, 798–800 (1987).
13. Marrack, P. *et al. J. exp. Med.* **158**, 1077–1091 (1983).
14. Saizawa, K., Rojo, J. & Janeway, C. A. *Nature* **328**, 260–263 (1987).
15. Janeway, C. A. & Katz, M. E. *J. Immun.* **134**, 2057–2063 (1985).
16. Dialynas, D. P. *et al. Immunol. Rev.* **74**, 29–56 (1983).
17. Golstein, P. *et al. Immunol. Rev.* **68**, 5–42 (1982).
18. Rits, M., Cormont, F., Bazin, H., Meykens, R. & Vaerman, J.-P. *J. immunol. Meth.* **89**, 81–87 (1986).
19. MacDonald, H. R. & Zaech, P. *Cytometry* **3**, 55–58 (1982).

Vaccinia virus encodes a secretory polypeptide structurally related to complement control proteins

Girish J. Kotwal & Bernard Moss

Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

Several polypeptides are secreted into the medium of cells infected with vaccinia virus, a cytoplasmic DNA virus belonging to the poxvirus family. One of these, a polypeptide of relative molecular mass 19,000¹ is structurally related to epidermal growth factor^{2–4} and binds to epidermal growth factor receptor stimulating proliferation of uninfected cells *in vitro*^{6–7} and *in vivo*⁸. Here, we show that a second, and much more abundant secretory polypeptide, is also encoded by vaccinia virus and is structurally related to the superfamily of complement control proteins. Members of this family can block complement-mediated induction of the inflammatory response, and engulfment, killing and lysis of bacteria and viruses^{9,10}.

Earlier studies¹¹ indicated that two major polypeptides of relative molecular mass (M_r) 35,000 (35K) and 12,000 (12K) are present in the medium of cells infected with vaccinia virus. We confirmed this, and discovered that neither polypeptide was secreted by cells infected with an attenuated mutant of vaccinia virus (designated 6/2)¹² which has a large deletion near the left end of the genome (Fig. 1). These results indicated that both secretory polypeptides were encoded in the region that had been deleted. The 35K protein was purified from the medium of cells infected with wild-type virus and the NH₂-terminal sequence was determined. This was compared with each of the 17 open reading frames, deduced from the nucleotide sequence of the DNA segment deleted in the mutant virus (G. Kotwal, unpublished), and one was found which matched exactly (Fig. 2).

This sequence started downstream of an initiating methionine indicating that during translation a signal peptide of 19 amino acids had been cleaved. The entire open reading frame, of 263 amino acids, has a calculated mass of 28.6K. The apparent over-estimation by polyacrylamide gel electrophoresis, of the

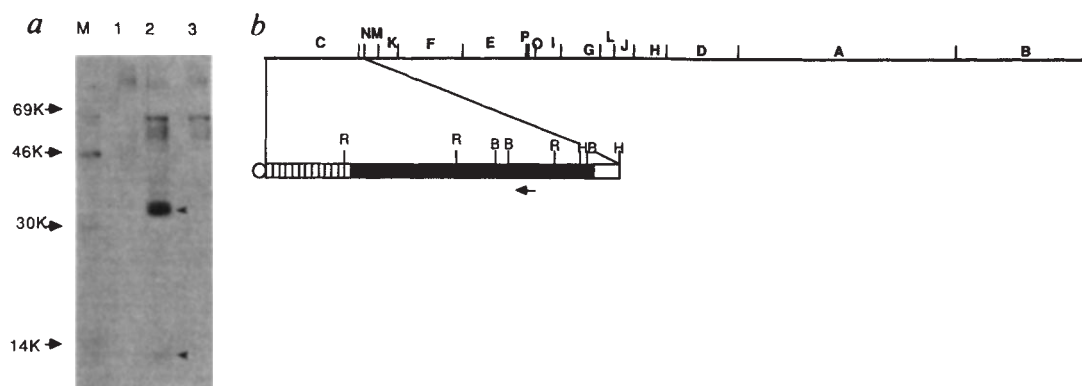
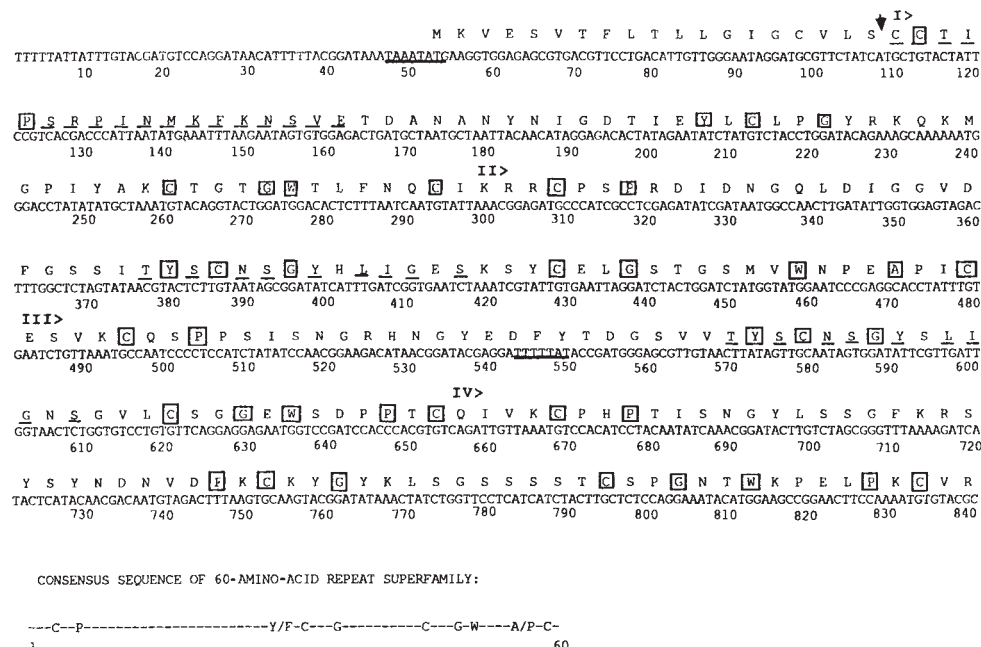


Fig. 1 Proteins secreted by cells infected with wild-type vaccinia virus or deletion mutant 6/2. *a*, An autoradiograph of a polyacrylamide gel is shown. Lanes are: M, ^{14}C -labelled relative molecular mass marker proteins; medium from rabbit kidney (RK-13) cells that were 1, uninfected; 2, infected with wild-type vaccinia virus; 3, infected with vaccinia virus deletion mutant 6/2. Arrows point to 35K and 12K polypeptides secreted only from cells infected with wild-type vaccinia virus. *b*, Representation of the *Hind*III map¹³ of the genome of vaccinia virus. The left end of the genome is enlarged; the terminal repeat region is indicated by vertical bars; the deleted region of 6/2 is indicated by the shaded area; R, B and H refer to *Eco*RI, *Bam*HI and *Hind*III sites, respectively. The arrow indicates the direction and position of the open reading frame corresponding to the 35K protein.

Methods. RK-13 cells were infected at a multiplicity of infection of 50 with either wild-type vaccinia virus (strain WR) or a deletion mutant of the WR strain (6/2) in minimal medium (RPMI). After 1 h, $50\ \mu\text{Ci ml}^{-1}$ of [^{35}S]methionine was added. The incubation was continued for 15 h, the medium harvested, and the secreted proteins prepared for SDS-polyacrylamide gel electrophoresis.

Fig. 2 Nucleotide sequence of the DNA including and flanking the gene encoding the 35K protein. The order of amino acids, represented in the one letter code, was deduced from the nucleotide sequence. The site of cleavage of the signal peptide (indicated by a downward arrow) was determined from the NH_2 -terminal sequence of the secreted protein. The underlined amino acids beginning with T indicate the location of a repeated sequence. The amino acids that are boxed conform to the consensus sequence of the 60-amino-acid repeat superfamily which is shown at the bottom of the figure. Roman numerals I, II, III and IV indicate the start of the four-tandem 60-amino-acid repeating units. Nucleotide sequences similar to a late transcription start site (TAAATATG)¹⁴ and an early transcription termination signal (T_3NT)¹⁵, that frequently occurs in the middle of late genes, are underlined.



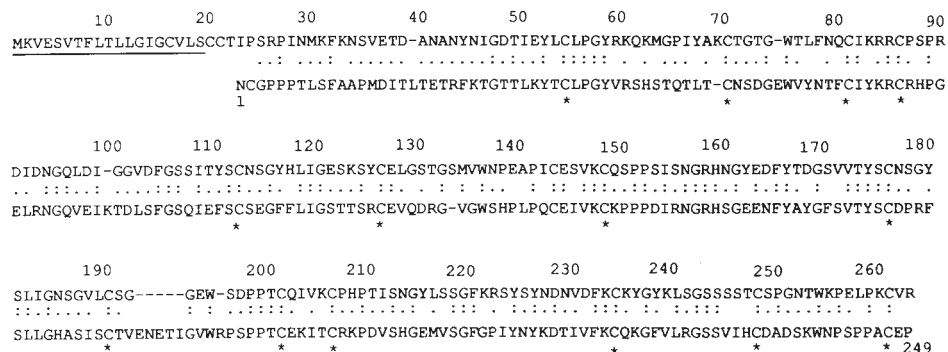
Methods. DNA fragments were cloned in M13 vectors, digested with exonuclease III to generate overlapping segments, and the nucleotide sequence of both DNA strands was determined by the dideoxy chain termination method. The secreted 35K protein was separated on a 12.5% SDS-polyacrylamide gel and transferred to an Immobilon filter. The NH_2 -terminal sequence was determined with an Applied Biosystems 477A gas-phase protein sequencing system. The result of the first 18 cycles was -, -, T, I, P, S, R, P, I, N, M, K, N, S, V, E. Because cysteines are not detectable by this method, their release in the first two cycles could only be inferred.

size of the processed polypeptide is not due to glycosylation because there are no glycosylation sites; moreover incorporation of glucosamine was not detected (G. Kotwal, unpublished). The deduced amino acid sequence of the secretory polypeptide revealed internal repetitions of which the most striking was Thr-Tyr-Ser-Cys-Asn-Ser-Gly-Tyr-His/Ser-Leu-Ile-Gly-Glu/Asn-Ser. This occurred between amino acids 90-103 and 154-167, numbered from the NH_2 terminus of the mature polypeptide (Fig. 2). In addition, conserved elements of the 60-amino-acid repeating unit of a superfamily of structurally related eukaryotic proteins^{9,10} are present in four places, comprising the entire secreted form of the vaccinia polypeptide (Fig. 2). This superfamily contains at least 6 complement control pro-

teins, of which 5 interact with C3b or C4b, and three non-complement proteins.

A search of the NBRF protein data base and GenBank revealed that the protein with the greatest similarity to the vaccinia virus 35K secretory protein is the human C4b-binding protein which contains eight 60-amino-acid repeating units¹⁶. The smaller vaccinia protein has a 38% identity with the first half of C4b (Fig. 3) and a 28% identity with the second half (data not shown). Moreover, in the homologous regions, 16 out of the 17 cysteines are conserved, as are the majority of prolines and glycines, indicating that the structural similarities may be even greater than the sequence ones (Fig. 3). Interestingly, the only gap of more than one amino acid which was inserted into

Fig. 3 Optimal alignment of the deduced amino acid sequence of the vaccinia virus open reading frame encoding the 35K secretory protein (top) with the human C4b-binding protein (bottom). The signal peptide, absent from the mature protein is underlined. Identical amino acids are connected by a colon and conservative amino acids by a full stop. Cysteine residues present in both proteins are highlighted by asterisks. Gaps imposed to maximize alignment are indicated by dashes. The sequence comparison was based on the FastP program of Lipman and Pearson¹⁷.



the vaccinia sequence to improve the alignment with the C4b-binding protein was opposite a potential glycosylation signal in the latter. Significant matches (27–28% identity) also occurred between the vaccinia secretory protein and the mouse complement factor H precursor, human beta-2-glycoprotein I, and complement factor B.

The complement system is composed of at least 20 plasma glycoproteins, of which seven have control functions. The C4b binding protein has a key role in regulating the classical complement pathway. It is remarkable that when the algorithm of Lipman and Pearson¹⁷ was used to compare the human C4b binding protein with the entire protein database, no entry had a matching score as high as that of the vaccinia virus secretory protein. The ability of a virus to negatively regulate the complement cascade could presumably help counteract host immune defences. Biochemical and genetic approaches are planned to determine the role of the vaccinia protein. Preliminary data indicate that the medium of cells infected with wild-type vaccinia virus, but not with the 6/2 deletion mutant, contains a factor that inhibits the classical pathway of complement-mediated haemolysis (G. Kotwal, unpublished).

It is generally appreciated that viruses regulate the metabolic machinery of the host cell to support their own replication. The discovery of the secretion by vaccinia virus of a growth factor^{5,6} that stimulates the proliferation of neighbouring cells *in vivo* prior to their infection⁸ indicated that there may be longer range extracellular viral effects. The virus-encoded secreted protein discussed here, that has structural similarity to the superfamily of complement control proteins extends this idea. It is possible that such 'virokines' represent an important class of proteins that have been largely overlooked because they are not essential for virus replication in tissue culture.

We thank Michael Raum and John Coligan for the protein sequence analysis and Patricia Earl for help with the amino acid homology searches.

Phenotypic heterogeneity of cerebrospinal fluid-derived HIV-specific and HLA-restricted cytotoxic T-cell clones

K. K. Sethi, H. Näher* & I. Stroehmann†

PROGEN Biotechnical Laboratories, Heidelberg, FRG

* Department of Dermatology, University of Heidelberg, Heidelberg, FRG

† Immunology Department, Medical Clinic, University of Bonn, Bonn, FRG

A variety of clinical syndromes, including AIDS and neurological disorders, may follow as a consequence of infection with the human immunodeficiency virus type 1 (HIV-1)^{1–4}. It is not yet clear, however, to what extent the destruction of lymphocytes and neural cells associated with these conditions is caused by adverse immune responses to HIV-1 or how much is due to cytopathic effects of the virus itself. Here we document the existence of HLA-restricted, HIV-1-specific cytotoxic T lymphocytes in the cerebrospinal fluid of two AIDS patients manifesting neurologic disorders. These cytotoxic T lymphocytes showed dual specificity⁵, recognizing target cells coated with purified HIV-1 envelope glycoprotein (gp 120) or inactivated HIV-1 in the context of HLA antigens. Cytotoxic T-cell clones derived from one of the AIDS patients revealed restriction specificities representing both HLA class I and HLA class II antigens. Considerable phenotypic heterogeneity was observed amongst these clones, some expressing conventional combinations of cytotoxic T-cell surface markers, and others displaying unusual phenotypes. The presence of HIV-specific cytotoxic T lymphocytes in AIDS patients, and in particular in their cerebrospinal fluid, suggests that these cytotoxic effectors may participate in the lymphoid cell and/or neurologic damage observed in such patients.

Cells derived from the cerebrospinal fluid (CSF) of two AIDS patients (M.T. and R.N.) with neurologic disorders were depleted of Fcγ-bearing cells, which exert conventional natural killer activity, and expanded *in vitro* before being assayed for cytotoxicity against a panel of target cells (Table 1). The CSF-derived cells (CSF-C) from each of the two patients efficiently lysed HIV-1-infected autologous macrophages and HIV-1- or gp120-coated target cells which were HLA-A, -B or -DR matched with respect to at least one locus. They did not, however, cause significant lysis of matched targets which were left uncoated or targets which were coated with herpes simplex virus (HSV) type 1. Nor did they lyse HIV-1 or gp120-coated target cells which were HLA-A, -B and -DR mismatched. Treatment of CSF-C with anti-CD3 monoclonal antibody and complement, which selectively depleted T cells, abolished this specific cytotoxicity confirming that the lytic CSF-C represented T cells (data not shown).

Received 13 June; accepted 28 July 1988.

- Venkatesan, S., Gershowitz, A. & Moss, B. *J. Virol.* **44**, 637–646 (1982).
- Brown, J. P., Twardzik, D. R., Marquardt, H. & Todaro, G. J. *Nature* **313**, 491–492 (1985).
- Blomquist, M. C., Hunt, L. T. & Barker, W. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7363–7367 (1984).
- Reisner, A. H. *Nature* **313**, 801–803 (1985).
- Stroobant, P. *et al.* *Cell* **42**, 383–393 (1985).
- Twardzik, D. R., Brown, J. P., Ranchalis, J. E., Todaro, G. J. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5300–5304 (1985).
- King, C. S., Cooper, J. A., Moss, B. & Twardzik, D. R. *Molec. cell. Biol.* **6**, 332–336 (1986).
- Buller, R. M. L., Chakrabarti, S., Moss, B. & Frederickson, T. *Virology* **164**, 182–192 (1988).
- Reid, K. B. M. *et al.* *Immun. Today* **7**, 230–234 (1986).
- Campbell, R. D., Law, S. K. A., Reid, K. B. M. & Sim, R. B. A. *Rev. Immun.* **6**, 161–195 (1988).
- McCrae, M. A. & Pennington, T. H. *J. Virol.* **28**, 828–834 (1978).
- Moss, B., Winters, E. & Cooper, J. A. *J. Virol.* **40**, 387–395 (1981).
- DeFilippes, F. M. *J. Virol.* **43**, 136–149 (1982).
- Rosel, J., Earl, P., Weir, J. P. & Moss, B. *J. Virol.* **60**, 436–449 (1986).
- Yuen, L. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6417–6421 (1987).
- Chung, L. P., Bentley, D. R. & Reid, K. B. M. *Biochem. J.* **230**, 133–141 (1985).
- Lipman, D. J. & Pearson, W. R. *Science* **227**, 1435–1441 (1985).

Table 1 HLA restriction of HIV-1-specific lysis mediated by CSF cells (CFS-C)

CSF-C from patient M.T. (HLA: A1, 3; B7, 18; Cw3, 6; DR2, 4)						CSF-C from patient R.N. (HLA: A2, w24; B12, 40; Cw6, —; DR1, 5)					
HLA antigens shared with		Targets infected/coated with:				HLA antigens shared with		Targets infected/coated with:			
targets:	—	HSV	HIV-1	gp120		targets:	—	HSV	HIV-1	gp120	
MT	Autologous	—2	NT	51	NT	Cw6	NT	NT	2.9	NT	
RN	Cw6	NT	NT	3	NT	Autologous	NT	NT	60	NT	
GS	A1; B7	—0.7	0.3	22	36	None	—1.5	4	6	0.5	
JE	B7; DR4	—0.3	3	39	29	A2, B40	—1.1	2	31	26	
SS	A3; B7; DR4	—1.2	0.9	31	26	A2	0.7	3	20	29	
BR	A1; B7; DR2, 4	—1.4	0.6	30	34	A2; B40	—1.5	4	26	30	
MS	DR2, 4	—0.9	0.4	14	19	A2	—0.9	5	31	29	
LT	B7	—1.9	2	31	36	A2; DR1	3	2	37	36	
HR	A1, DR2	—0.3	2	35	30	B40; DR5	—2.4	5	30	39	
DM	A3; B7, DR4	—1.7	0.8	32	38	A2; DR1	0.6	0.9	33	38	
PL	None	0.7	3	9	4	Aw24; DR1	—1.7	2	29	44	
BK	B7; Cw6; DR4	—1.9	9	30	35	A2; B40; Cw6	—1.0	8	33	37	
CEM	A1	—1.2	5	34	40	B40	0.9	3	29	34	
U937	A3; B18; Cw3	—0.9	2	5	29	None	—1.6	5	11	3	

To obtain CSF effectors, CSF (15 ml) from two AIDS patients (M.T. and R.N.) was added to growth medium (GM-HS) consisting of RPMI-1640, 10% pooled human AB serum (HIV-antibody negative), 2 mM glutamine antibiotics. Conventional NK cells were depleted by incubation of 10^5 cells with anti-CD16 (FcY receptor) monoclonal antibody and rabbit complement, and CSF-cells (10^4) were then cultured with 10^6 autologous irradiated (5,000 rads) peripheral blood mononuclear cells (PBMC), coated with 500- μ g lentil-lectin chromatography-purified HTLV-III gp 120 (ref. 16), in the presence of IL-2 (20% v/v). After expansion, bulk cultures were cryopreserved or cloned by limiting dilution technique. HIV-1-associated cytoplasmic antigen(s) in the CSF-C could not be detected by indirect immunofluorescence (IF) using HIV-1-positive serum and fluorescein isothiocyanate (FITC) conjugated goat anti-human Ig antiserum. As targets, human blood monocyte-derived macrophages, or human T leukemic (CEM) or monocytic (U937) cell lines were used. Macrophage cultures were prepared in Teflon vessels using GM-HS as described in ref. 17 and infected after seven days with HIV-1, CEM and cloned U937 cells were maintained as suspension cultures in GM supplemented with 10% fetal calf serum (GM-FCS). Before infection, 2×10^6 cells were pretreated with $2 \mu\text{g ml}^{-1}$ of polybrene (37 °C for 30 min) and incubated with 1 ml of a concentrated suspension ($10^{5.5} \times 50\%$ tissue culture infectious dose per ml) of UV-inactivated (Osram bulb; $2 \times 10^5 \text{ mJ cm}^{-2}$) HIV-1 (HTLV-III) for 2 h at 37 °C. Viral antigen was detectable in 60–90% of various targets as monitored by IF. Twelve-day-old macrophage cultures from the two AIDS patients 30–40% HIV-1-antigen-positive cells and were used as autologous infected targets following treatment with HTLV-III gp120 envelope protein (8 μg per 10^4 cells) for 4 h at 37 °C, which left >50% of cells positive for gp120 as monitored by IF. HSV-1 (strain thea) infected targets were prepared by adsorbing the virus (5 PFU per cell) at 37 °C for 1 h, as previously described^{18,19}. The labelling of target cells with $\text{Na}_2^{51}\text{CrO}_4$ and the ^{51}Cr release assays were carried out according to standard procedures described in detail previously^{19,20}. Briefly, aliquots (100 μl ; 10^4 cells) of test targets were mixed with 100- μl suspensions of effectors to give different effector to target (E:T) ratios. After 4-h incubation at 37 °C, 100- μl aliquots of supernatant were collected and counted in a gamma counter. Per cent specific ^{51}Cr release was calculated according to the formula $100 \times (E-S)/(M-S)$, where E = test ^{51}Cr release with effectors, S = spontaneous release in medium alone, M = maximum releasable ^{51}Cr with 1% Triton X-100. Spontaneous release from infected of uninfected targets ranged from 10–15% of maximum releasable counts. The tabulated values (means of triplicates) are of a typical experiment (E:T = 20:1). NT = not tested.

The observed susceptibility of targets prepared by exogenously binding HIV-1 gp120 or inactivated HIV-1 to cell surfaces suggests cytotoxic T lymphocytes (CTL) can recognize purified gp120 determinants in conjunction with HLA molecules and that CTL directed to gp120 determinants are generated during the host's immune response to HIV-1. The mechanism(s) by which exogenously bound gp120 is processed for CTL recognition in the context of HLA antigens requires investigation. CTL recognition after the addition of exogenous antigen has also been observed in the case of HLA class I restricted influenza virus-specific CTL, although in this case the antigen was added in the form of a small peptide⁶.

To investigate the HLA restriction specificities of the HIV-1-specific CTL more precisely, attempts were made to derive CTL clones from bulk cytolytic cultures originating from the CSF-C of patient M.T. and maintained in the presence of interleukin-2 (IL-2). A total of 55 cloned cultures were successfully expanded in a medium containing IL-2. Twenty-five of these 55 clones maintained homogeneous cell-surface phenotypes, and were examined for lytic activities. Natural killer (NK)-like activity, as monitored by lysis of K562 targets was exhibited by 10 of the 25 T-cell clones assayed (manuscript in preparation). Table 2 records the HIV-1-specific and HLA-restricted lytic effects of the remaining 15 cloned cultures. The four clones in Set I displayed unequivocal HIV-1-specific lysis in the context of HLA-class I determinants; three efficiently lysed HIV-1 treated targets with which they shared an HLA-A1 allele, the fourth,

MT 47 showed HIV-1-specific lysis against infected targets with which it shared HLA-B7, and all four failed to exert virus-specific lytic effects against other targets included in this panel. The phenotype of these four clones was CD3 (T3)⁺, CD8 (T8)⁺, CD4 (T4)[−]. In contrast, five cloned cultures demonstrated HIV-1-specific lysis which appeared to be restricted by HLA-class II determinants (set II). All five of these CTL clones showed specific lysis of HIV-1 infected targets which expressed DR2 molecules. Three of these five clones (MT.9, MT.33 and MT.67) were CD3⁺, CD4⁺, CD8[−] whereas the other two (MT.62; MT.129) coexpressed both CD4 and CD8 markers. Six individual clones categorized under set III consistently displayed specific lysis of HIV-1 infected targets expressing both B7 (HLA class I) and DR4 (HLA class II) antigens but not B7 or DR4 alone. Phenotyping of these six clones revealed two patterns: four of the clones coexpressed both CD4 and CD8 surface markers whereas the other two (MT.11 and MT.115) lacked both CD4 and CD8 markers.

Data from blocking experiments with monoclonal anti-HLA antibodies (Table 3) confirmed the HLA class I and class II restriction patterns of the CTL clones. The antibody W6/32, which detects a non-polymorphic HLA Class I epitope, but not the antibody OKIa1, which is directed against HLA-class II antigens, could reduce the lytic potential of each of the four HLA-class I restricted clones. Conversely, OKIa1 but not W6/32 antibody could block lysis mediated by each of the five HLA class II antigen-restricted CTL clones. The W6/32 antibody had

Table 2 HLA restriction patterns and cell-surface phenotypes of anti-HIV-1 CTL clones

CTL clones	Phenotype			Virus-specific lysis (%) of HIV-1-coated targets									
	CD3	CD4	CD8	GS	DM	HR	JE	MS	BK	CEM	BE	LT	SS
Set 1													
MT.12	+	−	+	61	−1	57	−1	−2	−1	51	50	2	−3
MT.25	+	−	+	25	−2	24	−2	2	2	30	28	1	−1
MT.71	+	−	+	43	3	47	−1	3	−1	51	39	−2	−2
MT.47	+	−	+	36	40	3	46	5	30	2	24	34	39
Set 2													
MT.9	+	+	−	−2	−1	53	−1	41	−2	2	30	−1	−1
MT.33	+	+	−	1	−2	37	−1	30	−1	3	26	−2	−1
MT.62	+	+	+	2	2	48	−3	52	−1	−1	37	−1	2
MT.67	+	+	−	3	1	29	−1	29	−1	−2	20	−1	3
MT.129	+	+	+	1	2	31	−2	43	−1	−2	29	−1	−1
Set 3													
MT.11	+	−	−	2	51	−1	44	3	38	−1	27	−2	36
MT.29	+	+	+	−1	39	−2	39	−2	41	−2	48	−1	47
MT.39	+	+	+	3	41	−4	60	1	36	−1	29	−1	38
MT.55	+	+	+	0	36	−4	29	0	41	−3	38	−1	33
MT.85	+	+	+	−3	29	−1	31	1	20	2	25	2	42
MT.115	+	−	−	2	43	−2	50	−1	29	−4	33	4	40
HLA antigens expressed by both target and effector cells				A1; B7	A3; B7; DR4	A1; DR2	B7; DR4	DR2, 4	B7; Cw6; DR4	A1	A1; B7; DR2, 4	B7	A3; B7; DR4

To obtain CTL clones, bulk T-cell cultures derived from the AIDS patient M.T. (HLA: A1, 3; B7, w50; Cw3, b; DR2,4) and were cloned by limiting dilution (0.3 to 1 cell per well) using 20% IL-2 containing medium in 96-well microtiter plates and 10^5 gp120 coated and irradiated autologous PBMC per well. Cells from wells showing growth were operationally designated as clones. They were expanded in culture flasks using medium containing IL-2 and were restimulated monthly. Cell-surface phenotypic analysis was performed by indirect IF using monoclonal antibodies: anti-OKT3, anti-OKT4 or anti-OKT8 (Ortho diagnostics, FRG) and FITC-conjugated F(ab)₂ fragments of anti-mouse IgG. Briefly, 10^4 cells were incubated with 100- μ l diluted (1:50) test antibody for 30 min at 4 °C, the cells washed and incubated with 200- μ l FITC-labelled anti-mouse IgG for 30 min at 4 °C. After rewashing, the cells were suspended in 200- μ l 1% paraformaldehyde in normal saline and examined under the fluorescence microscope. For double marking, 10^6 cells were incubated (45 min at 4 °C) with FITC-labelled anti-CD4 antibody followed by rhodamine-labelled anti-CD8 antibody (Sera-lab, UK). The cells were washed and incubated with Texas Red-avidin (Sera-lab; UK) and after rewashing suspended in 1% paraformaldehyde before examination. HIV-1-specific lysis was calculated by subtracting the % specific 51 Cr release for uninfected targets from the % specific 51 Cr release from appropriate (homologous) infected targets; for 51 Cr release assays; see legend to Table 1. The values for the various uninfected targets were very low and never exceeded 5% in various E:T combinations. In all cases dose-dependent release of 51 Cr was observed; values shown are for E:T ratio of 20:1.

Table 3 Effect of anti-HLA antibodies on HIV-1-specific lysis

CTL clones	HLA restriction specificity	Target cells	Virus-specific lysis (%) of HIV-1-infected targets in the presence of:		
			Medium	W6/32	OKIa1
MT.12	A1	CEM	30	3	34
MT.25	A1	CEM	35	9	27
MT.47	A1	CEM	47	12	37
MT.71	B7	BR	32	5	30
MT.9	DR2	BR	38	36	7
MT.33	DR2	BR	44	40	10
MT.62	DR2	BR	40	39	6
MT.67	DR2	BR	30	32	7
MT.129	DR2	BR	31	4	37
MT.11	B7/DR4	BR	37	10	46
MT.29	B7/DR4	BR	36	7	40
MT.39	B7/DR4	BR	25	9	20
MT.55	B7/DR4	BR	29	4	32
MT.85	B7/DR4	BR	33	10	40
MT.115	B7/DR4	BR	31	6	38

A monoclonal antibody directed against a non-polymorphic HLA class I epitope (W6/32; Serotech, England) and an anti-HLA class II Mab (OKIa1; Ortho Pharmaceuticals, USA) were tested for their ability to block lysis by CTL clones derived from the CSF of patient M.T. (HLA: A1; B7, 18; Cw3, 6; DR2, 4). 51 Cr-labelled target cells were incubated with W6/32 (1:200) or OKIa1 (1:100) (25 μ l per 10^4 cells) at 25 °C for 30 min before the addition of effectors. HIV-1-specific lysis was calculated by subtracting % specific 51 Cr release for uninfected targets from % specific 51 Cr release from appropriate (homologous) infected targets; for 51 Cr-release assays see legend to Tables 1 and 2.

significant inhibitory effect on lysis mediated by each of the six clones which could recognize HIV-1-infected targets carrying the B7 and DR4 specificities; no blocking effect was observed with OKIa1. The observed preference of six independent CTL clones for HIV-1 infected targets expressing both B7 and DR4 specificities was unexpected and intriguing. One explanation is that for effective recognition of HIV-infected cells, these CTL clones require expression of both B7 and DR4 molecules on target cell surfaces. Another possible interpretation is that these clones are restricted by a B7 or DR4 variant antigen present on B7/DR4 positive cells. Alternatively, it could be argued that these CTL recognize a virally modified HLA class II molecule (DR4) in the context of class I (B7) HLA antigen.

In most human viral and alloimmune systems, specific CTL are restricted by HLA class I (HLA-A, B) antigens and are normally of the CD4⁺, CD8⁺ phenotype⁷, although CD4⁺, CD8⁻ CTL directed towards HLA class I antigens have been reported⁸. The CD4⁺, CD8⁻ subset which is generally considered to perform helper/inducer functions⁹ has also been found to mediate measles virus-specific¹⁰ and HSV-specific¹¹ lytic effects which are HLA class II restricted. In the current study, HIV-1-specific CTL clones with restriction specificities, both for HLA class I and class II antigens and expressing the traditional phenotypes, CD4⁺, CD8⁺ and CD4⁺, CD8⁻, respectively, could be derived from the CSF of a patient with AIDS. Two virus-specific CTL clones seemingly restricted by HLA class I determinants expressed the inappropriate phenotypes CD3⁺, CD4⁺, CD8⁺. Surprisingly, among a series of six HIV-1-specific CTL clones with the unusual phenotypes CD3⁺, CD4⁺, CD8⁺ (4 clones) and CD3⁺, CD4⁺, CD8⁻ (2 clones), a discordance in HLA

restriction specificities was observed. Although the lytic activities of the two HIV-1 specific clones (MT 11 and MT 115) with the phenotype CD3⁺, CD4⁺, CD8⁺ were dictated by HLA antigens, it will be interesting to examine whether they express the T-cell receptor gamma gene products¹² reportedly displayed by T-cell clones with NK-like lytic activities¹³.

Two earlier reports have presented evidence showing the presence of HIV-1-specific CTL in the peripheral blood¹⁴ and in the broncho-alveolar lavage¹⁵ of HIV-infected persons. In the latter case virus-specific CTL displayed the CD3⁺, CD8⁺ phenotype and demonstrated restriction for class I HLA antigens, but in the case of the peripheral blood-derived CTL the precise phenotypic identity of the effectors and their HLA restriction specificities were not investigated. To our knowledge, the current study is the first to demonstrate the presence of HIV-1-specific CTL in the CSF of patients with AIDS and to document clonal heterogeneity, both in the cell-surface phenotypes and HLA-restriction elements of the effector CTL.

The presence of lytic effectors in the CSF of patients with AIDS where normally only rare lymphocytes are encountered suggests that such cells may contribute to the neurologic damage seen in such patients. The experimental approach using soluble gp120 or inactivated HIV-1-coated cells as *in vitro* targets for

HIV-1-specific CTL, may allow identification of potential gp120 epitope(s) involved in CTL recognition events and of the underlying mechanism(s) by which the putative epitope(s) is presented to the CTL for effective recognition in association with appropriate HLA molecule(s).

Received 25 April; accepted 15 July 1988.

1. Barré-Sinoussi *et al.* *Science* **220**, 868-871 (1983).
2. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497-500 (1984).
3. Levy, J. A. *et al.* *Science* **225**, 840-842 (1984).
4. Hoo, D. D., Pomerantz, R. J. & Kaplan, J. C. *New Engl. J. Med.* **317**, 278-286 (1987).
5. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51-177 (1979).
6. Taylor, P. M., Davey, J., Howland, K., Rothbard, J. B. & Askonas, B. A. *Immunogenetics* **26**, 267-272 (1987).
7. Biddison, W. E. *J. clin. Immun.* **2**, 1-18 (1982).
8. Flomenberg, N. *et al.* *Eur. J. Immun.* **13**, 905-911 (1983).
9. Reinherz, E. L. & Schlossman, S. F. *Cell* **19**, 821-827 (1980).
10. Jacobsen, S. *et al.* *J. Immun.* **133**, 754-757 (1984).
11. Yasukawa, M. & Zarling, J. M. *J. Immun.* **133**, 422-427 (1984).
12. Brenner, M. B. *et al.* *Nature* **322**, 145-149 (1986).
13. Moretta, L. *et al.* *Eur. J. Immun.* **17**, 1229-1234 (1987).
14. Walker, B. D. *et al.* *Nature* **328**, 345-348 (1987).
15. Plata, F. *et al.* *Nature* **328**, 348-351 (1987).
16. Lierly, H. K., Matthews, T. J., Langlois, A. J., Bolognesi, D. P. & Weinhold, K. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4601-4605 (1987).
17. Crowe, S., Mills, J. & McGrath, M. S. *AIDS Res. hum. Retroviruses* **3**, 135-143 (1987).
18. Sethi, K. K. & Brandis, H. Z. *Immun. Forsch.* **150**, 162-173 (1977).
19. Sethi, K. K., Stroehmann, I. & Brandis, H. *Nature* **286**, 718-720 (1980).
20. Sethi, K. K., Stroehmann, I. & Brandis, H. *Infect. Immunity* **36**, 657-661 (1982).

Immunodeficiency virus *rev* trans-activator modulates the expression of the viral regulatory genes

Michael H. Malim, Joachim Hauber, Randy Fenrick & Bryan R. Cullen*

Howard Hughes Medical Institute and Department of Medicine, Duke University Medical Center, Durham North Carolina 27710, USA

The pathogenic human retrovirus human immunodeficiency virus type 1 (HIV-1) encodes two *trans*-acting nuclear proteins, *tat* and *rev*, whose functional expression is essential for viral replication *in vitro*¹⁻⁶. The *tat* protein greatly enhances the expression of both structural and regulatory genes of HIV-1 (linked to the viral long-terminal-repeat promoter element)⁶⁻¹³, whereas the *rev* gene product (previously termed *art* or *trs*¹⁴) has only been shown to be required for the synthesis of structural proteins^{15,16}. Here, we demonstrate that *rev* also moderates the expression of regulatory genes of HIV-1. It decreases the expression of messenger RNAs that encode the full-length form of the viral *tat* gene product or the *rev* protein itself, and induces the synthesis of a previously unreported, truncated *tat* protein. These actions of *rev* are mediated by a dramatic shift in the ratio of spliced to unspliced cytoplasmic HIV-1 mRNA. Therefore *rev* not only activates the synthesis of the viral structural proteins, but also modulates the level and quality of HIV-1 regulatory gene expression.

Functional expression *in trans* of the HIV-1 *rev* protein is required for the synthesis of the viral *gag* and *env* structural gene products^{15,16}, but its site of action is controversial. It is possible that the *rev* protein is required *in trans* for the efficient translation of mRNAs encoding the viral structural proteins because *env* mRNA levels seem unaffected by *rev* expression^{16,17}. But there is also evidence that only the smaller, doubly spliced viral mRNA species were expressed in the absence of *rev*, suggesting that this protein might instead modulate the splicing of viral transcripts to permit the functional

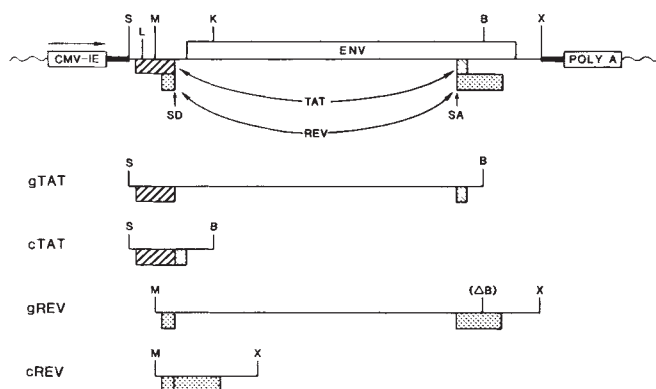


Fig. 1 Structure of the HIV-1 *tat* and *rev* gene expression vectors. A 3,108-base-pair (bp) *SalI* (S)—*XhoI* (X) genomic HIV-1 DNA fragment, containing the overlapping *tat* and *rev* genes as well as the complete *env* coding region, was excised from the replication-competent HXB-3 proviral HIV-1 clone^{18,19} and inserted into the expression vector pBC12/CMV (ref. 7). pgTAT was produced by removal of a 422-bp *BamHI* (B)—*XhoI* fragment containing part of the *rev* gene, preventing functional expression of *rev*. pgREV was similarly derived by deletion of a 170-bp *SalI*—*MstII* (M) fragment containing the *tat* gene initiation codon. pcTAT and pcREV include sequences derived from an HIV-1 cDNA clone⁵ and are exactly identical to their genomic equivalents pgTAT and pgREV, except that they lack the intron sequence shown. Thus, the mRNA encoded by pcTAT should be equal in size to the spliced mRNA encoded by pgTAT, at ~650 nucleotides whereas the unspliced mRNA encoded by pgTAT has a predicted size of ~3,000 nucleotides. The frame-shifted *rev* expression vector pgREVΔB was derived by filling in the unique *BamHI* site in pgREV with Klenow DNA polymerase. CMV-IE; cytomegalovirus immediate early promoter; SD, splice donor; SA, splice acceptor; L, *XhoII*; K, *KpnI*.

expression of the larger, singly (*env*) or unspliced (*gag*) viral structural gene mRNAs¹⁵.

Here, we have investigated the second hypothesis by examining the effect of *rev* on the expression of genomic and 'pre-spliced', complementary DNA forms of the *tat* gene (Fig. 1). The intron that coincides with most of the *env* gene coding region must remain unspliced for *env* mRNA expression, but this should prevent the fusion of the two *tat* coding exons. There

* To whom correspondence should be addressed.

Fig. 2 Immunoprecipitation analysis of transfected COS cell cultures. The HIV-1 replication-competent COS cell line²⁴ was transfected²⁵ with equal levels of the *tat* and *rev* expression vectors as shown. When only one plasmid is listed, the parental vector, pBC12/CMV⁷, was also co-transfected as a negative control. At 60 h after transfection the cells were labelled for 2 h with [³⁵S]cysteine (a) or [¹⁴C]leucine (b) and subjected to immunoprecipitation²⁵ using anti-*tat* (a) or anti-*rev* (b) specific peptide antisera^{5,6}. Precipitated proteins were resolved on 14% discontinuous SDS-polyacrylamide gels and visualized by autoradiography. aa, Amino acid.

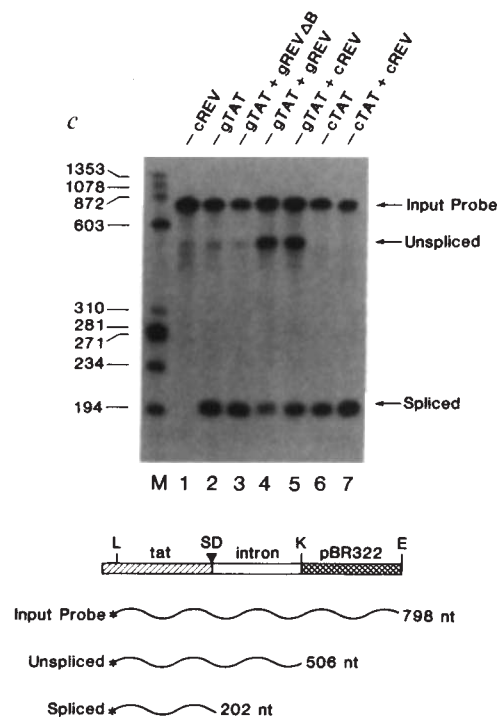
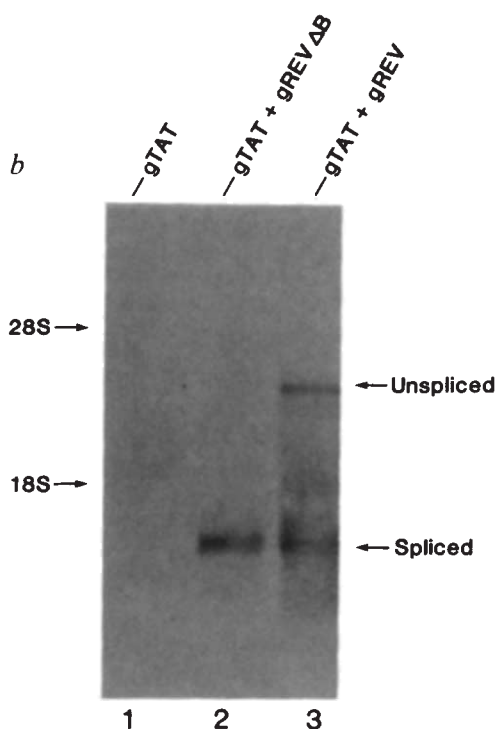
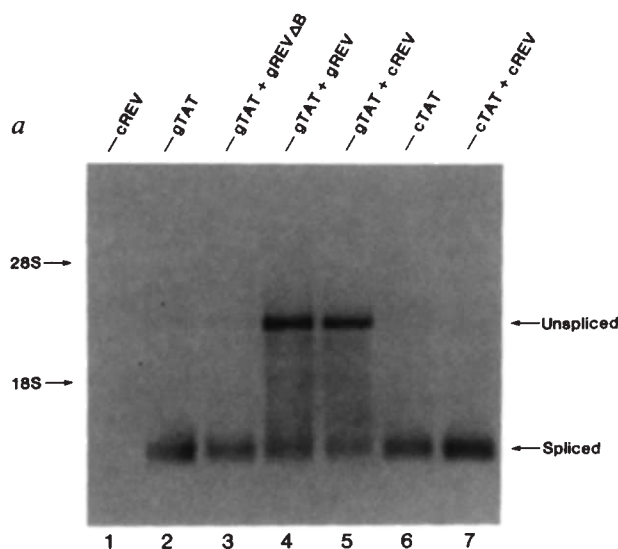
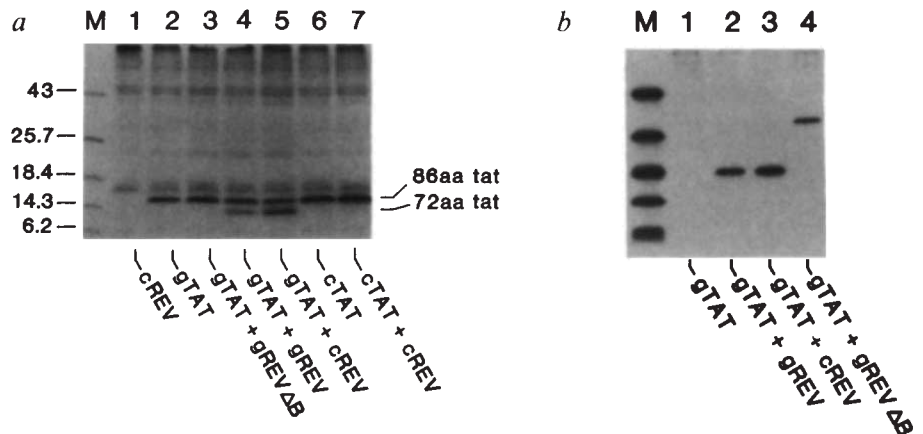


Fig. 3 Analysis of HIV-1 derived mRNA species expressed in transfected COS cell cultures. a, Northern analysis²⁶ of total cytoplasmic RNA using a *tat*-specific DNA probe derived from the 170-bp *Sal*I-*Mst*II fragment missing in the *rev* expression vectors. RNA was collected²⁷ 60 h after transfection and 4 µg aliquots used for each lane. A parallel lane was stained to localize the rRNA markers. b, Northern analysis²⁶ of total cytoplasmic RNA using a *rev*-specific DNA probe derived from the 422-bp *Bam*HI-*Xho*I fragment missing in the *tat* expression vectors. nt, nucleotide. c, Quantitative S1 nuclease protection analysis⁷ of the cytoplasmic RNA samples in Fig. 3a. The probe used was end-labelled⁷ at an *Xho*II (L) site within the *tat* coding region (Fig. 1) and extends through the predicted splice donor (SD) sequence into the *env* intron element. A pBR322 DNA 'tag' was attached at a *Kpn*I (K) site within the intron to allow us to distinguish the full-length, input probe from the probe fragments rescued by the unspliced and spliced transcripts, as indicated. Size markers were obtained by end-labelling a *Hae*III digest of phage φX174 replicative form DNA. 4 µg RNA were used for each lane. E, *Eco*RI.

is a translation termination codon next to the first *tat* coding exon, so a singly spliced form of *tat* mRNA should produce a truncated form of the *tat* protein^{7,12,13}. This additional translation termination codon is completely conserved in all strains of primate immunodeficiency virus so far examined, suggesting that it is functionally important¹⁸⁻²¹.

We have previously shown that the full-length 86-amino-acid form of the *tat* protein migrates anomalously slowly, with an apparent relative molecular mass (M_r) of ~15,500 (15.5K), whereas the single-exon 72-amino-acid form migrates with an M_r of ~14,000 (14K) (ref. 22). We therefore examined whether co-expression of *rev* had any effect on the relative mobility of the *tat* proteins encoded by genomic (pgTAT) or cDNA (pcTAT) *tat* expression vectors (Fig. 2a). Cells transfected with the pgTAT vector expressed one *tat* protein species with the mobility of the full-length form (Fig. 2a, lane 2); co-expression of *rev* resulted in the appearance of the truncated, 14K form of the *tat* protein and a reduction in levels of the full-length form (Fig. 2a, lanes 4 and 5). This effect was particularly marked when the co-transfected *rev* vector encoded a cDNA-derived (pcREV), rather than genomic (pgREV), *rev* mRNA, and was abolished by the introduction of a frame-shift mutation into the *rev* coding region (Fig. 2a, lane 3). However, cells transfected with the pcTAT vector, yielded only the full-length form of *tat*, regardless of *rev* co-expression (Fig. 2a, lanes 6 and 7).

Expression of the *rev* protein in the transfected cultures was also monitored by immunoprecipitation⁵. No signal specific to *rev* was observed in cells transfected with pgTAT alone (Fig. 2b, lane 1), but the *rev* protein (M_r ~19K) was readily detected in cultures transfected with either the pgREV or pcREV expression vectors (Fig. 2b, lanes 2 and 3). Cultures transfected with the pcREV vector consistently contained about three times as much *rev* protein as those transfected with the pgREV vector. A *rev* protein of aberrant mobility was detected in cultures transfected with the frame-shifted *rev* expression vector pgREVΔB (Fig. 2b, lane 4).

Total cytoplasmic RNA collected from the transfected cell cultures was subjected to Northern analysis using probes for *tat* or *rev* mRNAs (Fig. 3). Cultures transfected with pgTAT expressed only a single *tat*-specific transcript of the size predicted for a spliced *tat* mRNA (Fig. 1). As expected, this displayed the same mobility as the *tat* mRNA expressed in the pcTAT transfected cultures (Fig. 3a, lanes 2, 3, 6 and 7). Co-expression of *rev* in the culture transfected with pgTAT resulted in the appearance of high levels of an mRNA with the mobility predicted for the unspliced *tat* transcript (Fig. 3a, lanes 4 and 5). There was no equivalent band in cultures transfected with pcTAT and pcREV, nor was any signal specific for the *tat* gene observed in a culture transfected with pcREV alone.

Northern analysis using a *rev*-specific probe revealed that the *rev* protein also modulated the expression of its own mRNA transcript. A single species of mRNA of the mobility predicted for the spliced *rev* transcript was observed in cultures transfected

with the defective *rev* expression vector pgREVΔB (Fig. 3b, lane 2). But high levels of an mRNA of the mobility predicted for the unspliced transcript were observed in the pgREV transfected cells (Fig. 3b, lane 3). This resulted in a reduced level of spliced mRNA, which could explain the lower level of *rev* protein encoded by the pgREV vector compared with pcREV (Fig. 2b).

The same cytoplasmic RNA samples were analysed by quantitative S1 nuclease protection studies using a *tat* probe which crosses the predicted splice donor site (Fig. 3c). The results confirmed that *rev* co-expression is required for the stable expression of the unspliced mRNA encoded by pgTAT (Fig. 3c, lanes 2, 3 versus 4, 5). These results, together with additional data obtained using probes that includes the 3' portion of the pgTAT construction (data not shown), also demonstrated that the splice sites used by the transcripts encoded by the genomic expression vectors pgTAT and pgREV were identical to the sites used during the normal splicing of viral mRNA within HIV-1 infected cells²³.

We have examined the effect of *rev* on the expression of HIV-1 mRNAs that span the *env* gene coding region, and shown that unspliced mRNAs are functionally expressed only in the presence of the viral *rev* protein. We do not know how *rev* exerts this effect. The appearance of the shorter form of *tat* encoded by an unspliced HIV-1 transcript corresponds to the disappearance of the longer form of *tat* encoded by the spliced transcript when assayed at either the protein or mRNA level (Figs 2a and 3c). These observations are consistent with the hypothesis that the nuclear *rev* protein can modulate the processing of mRNA^{5,15} and indicate that the induced expression of the unspliced or singly spliced mRNAs that encode *gag* and *env* may be at the expense of the pool of fully spliced mRNAs which encode the HIV-1 regulatory proteins. But alternative hypotheses, such as a role for *rev* in the nuclear export of unspliced HIV-1 mRNAs, cannot be eliminated. Such a function could explain why *env* mRNAs that are unable to splice are poorly expressed in the absence of *rev*¹⁷.

The data presented here predict the existence of two forms of the HIV-1 *tat* protein. We have not as yet confirmed these observations in HIV-1 infected, CD4-bearing human T cells, but cDNA clones with the structure predicted for transcripts encoding the 72- and 86-amino-acid forms of *tat* have been isolated from HIV-1 infected cells²³. Current evidence suggests that both forms of the *tat* protein are equally able to activate HIV-1 gene expression driven by the long terminal repeats in transfected fibroblast cell lines^{7,12,13}. These two *tat* proteins could display significantly different properties within the infected lymphoid cell, however. The evidence presented here for their differential modulation by *rev* therefore raises the interesting possibility that this second *trans*-acting HIV-1 protein may affect both the quality and, indirectly, the quantity of HIV-1 specific gene expression.

We thank Sharon Goodwin for excellent secretarial assistance.

Received April 27; accepted 22 July 1988.

1. Fisher, A. G., Aldovini, A., Debouk, C., Gallo, R. C. & Wong-Staal, F. *Nature* **320**, 367-371 (1986).
2. Dayton, A. I., Sodroski, J. G., Rosen, C. A., Goh, W. C. & Haseltine, W. A. *Cell* **44**, 941-947 (1986).
3. Terwilliger, E. *et al.* *J. Virol.* **62**, 655-658 (1988).
4. Sadaie, M. R., Benter, T. & Wong-Staal, F. *Science* **239**, 910-914 (1988).
5. Cullen, B. R. *et al.* *J. Virol.* **62**, 2498-2501 (1988).
6. Hauber, J., Perkins, A., Heimer, E. P. & Cullen, B. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6364-6368 (1987).
7. Cullen, B. R. *Cell* **46**, 973-982 (1986).
8. Arya, S. K., Guo, C., Josephs, S. F. & Wong-Staal, F. *Science* **229**, 69-73 (1985).
9. Rosen, C. A. *et al.* *Nature* **319**, 555-559 (1986).
10. Wright, C. M., Felber, B. K., Paskalis, H. & Pavlakis, G. N. *Science* **234**, 988-992 (1986).
11. Peterlin, B. M., Luciw, P. A., Barr, P. J. & Walker, M. D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9734-9738 (1986).
12. Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F. & Haseltine, W. *Science* **229**, 74-77 (1985).

13. Muesing, M. A., Smith, D. H. & Capon, D. J. *Cell* **48**, 691-701 (1987).
14. Gallo, R., Wong-Staal, F., Montagnier, L., Haseltine, W. & Yoshida, M. *Nature* **333**, 504 (1988).
15. Feinberg, M. B., Jarrett, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. *Cell* **807**-817 (1986).
16. Sodroski, J. *et al.* *Nature* **321**, 412-517 (1986).
17. Knight, D. M., Flomerfelt, F. A. & Ghayeb, J. *Science* **236**, 837-840 (1987).
18. Shaw, G. M. *et al.* *Science* **226**, 1165-1171 (1984).
19. Ratner, L. *et al.* *AIDS Res.* **3**, 57-69 (1987).
20. Guyader, M. *et al.* *Nature* **326**, 662-669 (1987).
21. Hirsch, V., Riedel, N. & Mullins, J. I. *Cell* **49**, 307-319 (1987).
22. Hauber, J. & Cullen, B. R. *UCLA Symp. Molec. cell. Biol. New Ser.* **67**, 191-200 (1987).
23. Muesing, M. A. *et al.* *Nature* **313**, 450-458 (1985).
24. Levy, J. A., Cheng-Mayer, C., Dina, D. & Luciw, P. A. *Science* **232**, 998-1001 (1986).
25. Cullen, B. R. *Meth. Enzym.* **152**, 684-704 (1987).
26. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).
27. Greenberg, M. E. & Ziff, E. B. *Nature* **311**, 433-438 (1984).

Genome linking with yeast artificial chromosomes

Alan Coulson*, Robert Waterston†, Jane Kiff†, John Sulston* & Yuji Kohara*

* MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

† Department of Genetics, Washington University Medical School, St Louis, Missouri 63110, USA

The haploid genome of *Caenorhabditis elegans* consists of some 80×10^6 base pairs of DNA contained in six chromosomes¹. The large number of interesting loci that have been recognized by mutation², and the accuracy of the genetic map³, mean that a physical map of the genome is highly desirable, because it will facilitate the molecular cloning of chosen loci. The first steps towards such a map used a fingerprinting method to link cosmid clones together⁴. This approach reached its practical limit last year, when 90–95% of the genome had been cloned into 17,500 cosmids assembled into some 700 clusters (contigs), but the linking clones needed were either non-existent or extremely rare. Anticipating this, we had planned to link by physical means—probably by hybridization to *NotI* fragments separated by pulse field gel electrophoresis^{5–7}. *NotI* recognizes an eight base sequence of GC pairs; thus the fragments should be large enough to bridge regions that clone poorly in cosmids, and, with no selective step involved, would necessarily be fully representative. However, with the availability of a yeast artificial chromosome (YAC) vector⁸, we decided to use this alternative source of large DNA fragments to obtain linkage. The technique involves the ligation of large (50–1,000 kilobase) genomic fragments into a vector that provides centromeric, telomeric and selective functions; the constructs are then introduced into *Saccharomyces cerevisiae*, and replicate in the same manner as the host chromosomes.

Initially, we attempted to apply the same fingerprinting method used for the cosmids to the YACs directly, but were unable to obtain satisfactory data. With care, patterns of bands that give unequivocal matches^{9,10} can be produced from the smaller YACs (<150 kb), but the quality is always poor: the

bands tend to be doubled, the background is high, and discrimination between genuine bands and spurious bands (perhaps generated by partial digestion) is difficult. Reasoning that even if we were able to improve the quality of the data, we would still be restricted to the smaller YACs which are less useful for linking, we turned to hybridization as an alternative strategy.

The application of hybridization has proved to be quite straightforward, mainly because we had fortunately cloned about half of our previous genomic fragments into the Lorist series of vectors^{11,12}. These vectors have no appreciable homology with the YAC vector, so that direct hybridization of Lorist cosmids to YACs and *vice versa* is possible without first separating probes from vector sequences. We have presumably also benefited from the relatively low proportion (17%)¹ of repeated sequences in the nematode genome. Our strategy was to prepare two sets of ordered grids: one of random YAC clones, representing about 2.5 genome equivalents, and the other of cosmids, which we made as fully representative as possible of our contigs and unattached clones. By probing the YAC grid with cosmids from the ends of contigs, walks can be initiated (Fig. 1). The YACs thus selected can be separated from the host chromosomes (by pulsed field gel electrophoresis) and used to probe the cosmid grid (Fig. 2). The latter type of probing is particularly powerful, because it not only establishes linkage but also reveals the precise position and extent of the YAC with respect to the cosmid clones (Fig. 3).

The benefits of this approach have been immediate. In seven months, we have performed about 1,000 probings with both selected and randomly chosen YACs of up to 600 kb; as a result, the number of contigs has been reduced from 700 to 346. The YACs not only are larger than other sorts of clones but also (as suggested by Burke *et al.*⁸) show every indication of sampling sequences that are poorly represented in cosmid banks.

An important, indeed vital, step when hybridization is used to make joins is validation by some other method. Because there are dispersed repeated sequences in the genome, no single hybridization can be relied upon to indicate linkage. Also, the fidelity of the YAC cloning process is not yet established, so that clones could, in principle, contain sequences fused together which are not adjacent in the genome.

The most convincing verification of linkage is by restriction fragment size. In about one-third of the joins made, inspection

Fig. 1 Hybridization of an entire Lorist cosmid (T15G10, compare with Fig. 3) to a grid of YACs. Four strongly hybridizing colonies are present; the remainder of the clones show only background levels, which facilitates assignment of grid positions of the positive clones.

Methods. The YAC library was prepared with minor adaptations as described^{8,15}. The *C. elegans* DNA was prepared by extensive proteolytic digestion of starved, frozen L1 stage animals, followed by sedimentation through a 15–50% sucrose gradient. After partial *EcoRI* digestion, the DNA was size-fractionated over sucrose and fractions containing DNA fragments greater than 100 kb or 200–700 kb were pooled, concentrated and ligated to a 100-fold molar excess of the restricted, phosphatased pYAC4 vector. YAC colonies were initially selected in soft-agar on regeneration plates lacking uracil; individual colonies were picked onto complete plates lacking uracil, tryptophan and arginine, and supplemented with canavanine. Positive clones were grown in complete liquid medium lacking uracil in Micronic 1 ml tubes and frozen in 15% glycerol for long term storage. For hybridization, portions of colonies were transferred by toothpick to chosen grid positions on a complete plate lacking uracil. After overnight growth at 30 °C, the grid was lifted on to dry Hybond-N. Each of these masters was used to generate 70 or more replicas as follows. A sandwich was prepared consisting of (from bottom to top) a glass plate, a dry piece of Whatman 3MM paper, the master face up, a Hybond-N filter previously wetted face down on a complete plate lacking uracil, a piece of velvet and a second glass plate. Heavy hand pressure was applied briefly and the stack was dismantled, the copy and the master being placed face up on their respective plates. The cycle was repeated immediately, regrowth being neither necessary nor desirable. Masters could be kept for reuse by impregnation on agar containing 20% glycerol, and storage in dry dishes at –70 °C. The copies were grown at 30 °C for 20 h, laid briefly on Whatman 3MM paper soaked in 0.8% dithiothreitol in SOE (1 M sorbitol, 20 mM EDTA, 10 mM tris-Ac pH 8), and incubated overnight at 37 °C on circles of 3 MM soaked in 1.5 mg ml^{–1} zymolyase 20T (Seikagaku Kogyo Co., Tokyo), 1% 2-mercaptoethanol, SOE in dishes in sealed bags. The filters were denatured¹⁶, air dried, and UV-irradiated. Cosmid DNA was prepared by a miniprep procedure¹⁷, and oligolabelled^{18,19}. Conventional hybridization provided a satisfactory signal after a 6–24 h exposure to film with an intensifying screen.

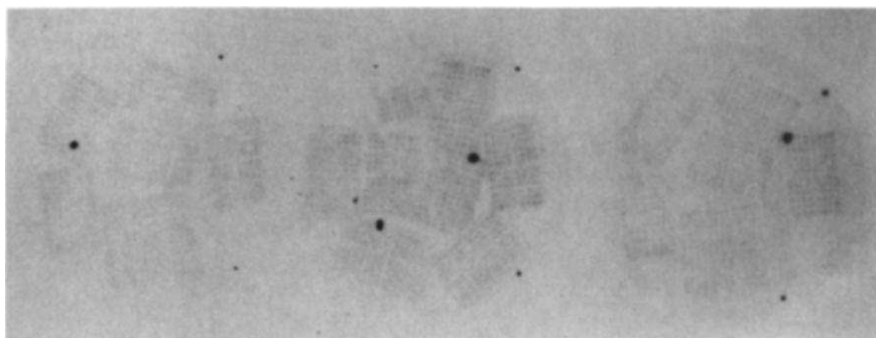


Fig. 2 Hybridization of an entire YAC (Y19C10, compare with Fig. 3) to a grid of Lorist cosmids. Grids of the stored cosmids were prepared in the same way as for YACs (Fig. 1) with the differences that the master filter material was BA85 nitrocellulose, the zymolyase step was omitted, and growth was at 37 °C on kanamycin-TYE. YAC probes were prepared as follows (adapted from the method in ref. 20). A patch of cells 1–2 cm² in area for each clone was grown on a complete plate lacking uracil for 48 h at 30 °C. Ninety-six such heavy streaks were collected by wire loops and stirred into 200 µl of 50 mM EDTA in round-bottomed microtitre wells. The dishes were centrifuged on styrofoam pads at 3,500 r.p.m. for 2 min, the supernatants were aspirated, and the pellets softened by vigorous vortexing. To each well was added 10 µl 50 mM EDTA, followed by 6 µl solution I (1 mg ml⁻¹ zymolyase 100T, 5% 2-mercaptoethanol, in SCE²¹) and, after 5 min, 40 µl 1% low gelling temperature agarose held at 42 °C in 0.1 M EDTA. The contents of the wells were immediately pipetted into a perspex mould containing 3.5 × 1.8 × 6 mm wells (arranged in the format of a 96-well microtitre dish) sealed at the bottom with adhesive tape. After hardening, the plugs were pushed out into 100 µl of solution II (7.5% 2-mercaptoethanol, 0.5 M EDTA-pH 9) in flat-bottomed microtitre wells. The plugs were then processed as described²⁰. For recovery of the YAC DNA the plugs were slipped into 3.8 × 1.5 × 7 mm wells in 0.7% low gelling temperature agarose in 0.5 × TBE¹⁶ gel and subjected to pulsed field gel electrophoresis in a CHEF apparatus⁶ at 5V cm⁻¹ with a 1-min switching interval. The gel was stained with ethidium bromide in water, then destained in 0.5 × TAE¹⁶. The YAC bands were cut out, melted at 70 °C, and 8 µl portions were oligolabelled^{18,19}.

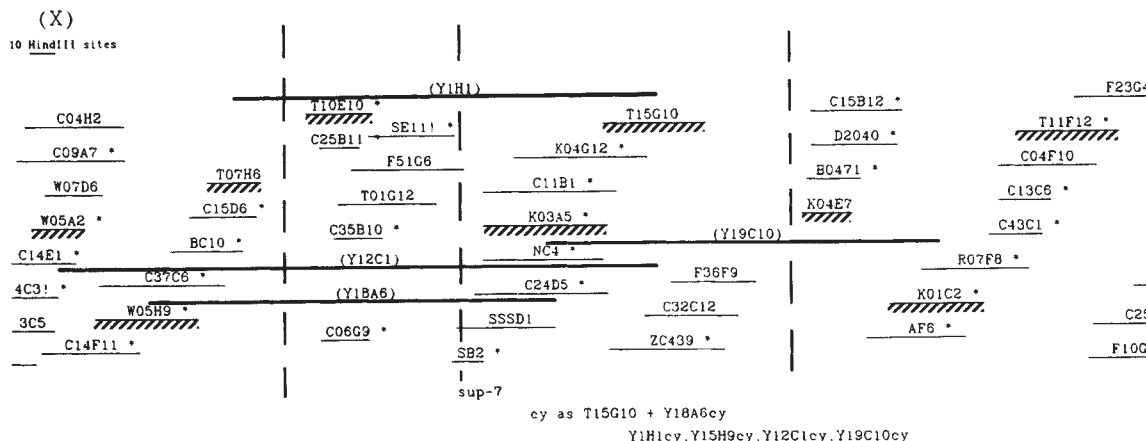
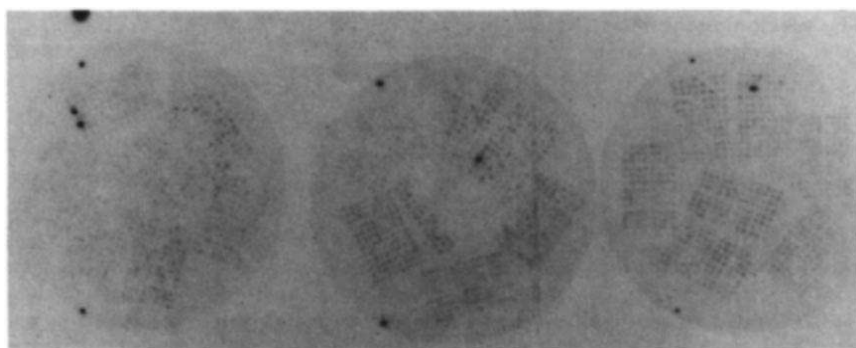


Fig. 3 An example of linkage by nested YACs. Previous cosmid-cosmid matching had established the four contigs delineated by vertical dashed lines. Cosmids (containing 35–40 kb inserts) are indicated by light horizontal lines, of lengths proportional to number of *Hind*III sites (see scale to left); asterisks indicate presence of additional, similar, clones in the database. Probing of the YAC grid with K03A5 resulted in the selection of four YACs (indicated by heavy horizontal lines; estimated insert sizes: Y1H1 230 kb, Y12C1 290 kb, Y18A6 180 kb, Y19C10 170 kb). The positions of these YACs, and, consequently, the linkage of the cosmid contigs, were established by probing of the cosmid grid (in which these contigs are represented by the cosmids indicated by hatched lines); the pattern of hybridization seen for Y19C10 is illustrated in Fig. 2. Three sorts of join are apparent. The centre one is known to be correct because of a small revealed overlap between F51G6 and SB2. The left hand one is strongly confirmed by the logical nesting of the YACs to multiple adjacent cosmids. The right hand one is not yet confirmed, because it might have resulted from a fusion of unrelated sequences in the cloning of Y19C10; note, however, that the hybridization to two cosmids in each contig indicates that we are not being misled by dispersed repeats. For the time being, the gaps between cosmid contigs have been arbitrarily set at five *Hind*III sites.

of the ends of two contigs that hybridize to the same YAC reveals a previously undetected overlap in the cosmid fingerprint data (Fig. 3). Frequently the overlap is no more than two bands; this is not enough for our computer matching techniques to find, but, provided the logical fit of the remaining bands with neighbouring cosmids is satisfactory, even two bands are powerful confirmation of the accuracy of the join. Occasionally, larger overlaps are found, which should ideally have been detected but due to poor data were not found by the matching routines. In all these cases visual inspection of the fingerprints allows unequivocal joining once the matching pair has been identified by the results of YAC hybridization.

The two-thirds of joins based solely on hybridization will continue to be doubtful until a later stage of the project. However, experience already indicates to us that the great majority of such joins are in fact genuine. In several instances two or more YACs bridge the same gap, and in the case of different but overlapping YACs we can also test for a logical fit to the structure of the pre-existing cosmid contigs (Fig. 3, for example). In other instances the hybridization of the YAC to more than

one non-overlapping cosmid in a contig indicates that we are not being misled by dispersed repeats. Linkage of genetically marked contigs with the YACs has been consistent with their known positions^{3,13}: two particular examples are linkage of *lin-14* to genetically adjacent dimorphisms (G. Ruvkun *et al.*, submitted), and a 2,200 kb contig on chromosome III that includes several cloned genes. Finally there have been no obvious paradoxes involving YACs well positioned within contigs, also indicating that false ligation events in the preparation of the library were rare.

As the linkage develops, paradoxes will reveal most of the erroneous joins. Nevertheless, we shall eventually want to confirm that all the surviving joins are correct. The possibility of repetitious artefacts can be eliminated by showing that end cosmids and joining YACs have fragments in common. Both repetition and false ligation can be ruled out by demonstrating the presence of the same bridging macrorestriction fragment in the YAC and the genome. A third approach, which is more laborious but ultimately more useful, is to obtain joining cosmids or λ clones by probing with the YAC and/or the end clones,

or by subcloning the YAC. The virtue of initially proceeding to full linkage at a more limited level of confidence, is that we and others can then concentrate on those areas of the map that are most urgently needed for molecular genetics.

In future, YACs are likely to be incorporated in mapping strategies from the outset. However, it is unlikely that YAC maps will supplant cosmid and λ clone maps, because the smaller clones are needed for practical uses such as localization of genes, transformation¹⁴, and sequencing. Whether it is more efficient to begin by making a YAC map and subsequently patch in the smaller clones, or to assemble contigs of small clones and then link them with YACs (as we, for historical reasons, have done), will have to be decided by experiment. The higher proportion of repeated sequences in more complex genomes may make our hybridization approach impracticable; its usefulness in such cases will be strongly dependent on the size and distribution of the repeats within the genome. Genomes that overall have a higher proportion of repeated sequences will not necessarily be more resistant to the approach. Nevertheless, even in the nematode a tenth of YAC probes yield unusable hybridization patterns. Thus, the development of fingerprinting procedures that can handle YACs directly may be crucial.

We thank David Burke for assistance with library construction, Maynard Olson, Georges Carle and David Burke for communication of methods and ideas, Donna Albertson and Rita Fishpool for *in situ* hybridizations, Humaira Ameer for technical assistance and the *C. elegans* community for unstinting contributions of genetically-marked clones to the map. R.H.W. is supported by USPHS.

Received 13 May; accepted 27 July 1988.

1. Sulston, J. E. & Brenner, S. *Genetics* **77**, 95–104 (1973).
2. *The nematode Caenorhabditis elegans* (ed. Wood, W. B., Cold Spring Harbor Laboratory, New York, 1988).
3. Edgley, M. L. & Riddle, D. L. *The nematode C. elegans*. Genetic Maps **4**, 351–365 (1987).
4. Coulson, A., Sulston, J., Brenner, S. & Karn, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7821–7825 (1986).
5. Schwartz, D. C. & Cantor, C. R. *Cell* **37**, 67–75 (1984).
6. Chu, G., Vollrath, D. & Davis, R. W. *Science* **234**, 1582–1585 (1986).
7. Southern, E. M., Anand, R., Brown, W. R. A. & Fletcher, D. S. *Nucleic Acids Res.* **15**, 5925–5943 (1987).
8. Burke, D. T., Carle, F. G. & Olson, M. V. *Science* **236**, 806–812 (1987).
9. Coulson, A. & Sulston, J. in *Practical Approaches* (ed. K. Davies, in press).
10. Sulston, J. *et al.* *CABIOS* **4**, 125–132 (1988).
11. Gibson, T., Coulson, A., Sulston, J. & Little, P. F. R. *Gene* **53**, 275–281 (1987).
12. Gibson, T., Rosenthal, A. & Waterston, R. *Gene* **53**, 283–286 (1987).
13. Albertson, D. G. *EMBO J.* **4**, 2493–2498 (1985).
14. Fire, A. *EMBO J.* **5**, 2673–2680 (1986).
15. Burke, D. T. thesis, Univ. Washington (1988).
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
17. Birnboim, H. C. & Doly, J. *Nucleic Acids Res.* **7**, 1513–1523 (1979).
18. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6–13 (1983).
19. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **137**, 266–267 (1984).
20. Carle, F. G. & Olson, M. V. *Meth. Enzym.* **155**, 468–482.
21. Olson, M. V., Loughney, K. & Hall, B. D. *J. molec. Biol.* **132**, 387–410 (1972).

Site-specific cleavage of single-stranded DNAs at unique sites by a copper-dependent redox reaction

Sergey A. Kazakov, Tatiana G. Astashkina,
Sergey V. Mamaev & Valentin V. Vlassov

Institute of Bioorganic Chemistry, Siberian Division of the USSR
Academy of Sciences, Lavrentiev Prospekt 8,
Novosibirsk 630090, USSR

Metal ions play a crucial role not only in the formation and maintenance of nucleic acid structure, but also in important biochemical conversions of polynucleotides. Some aqueous metal ions, acting as general acid/base (or electrophilic/nucleophilic) catalysts, can induce site-specific cleavage of RNA^{1–6}. DNA is not cleaved efficiently by the non-redox metal-induced mechanism⁷.

```

50
GATCC GTCGA GCTGC AGGGG GGGGG GGGGG GGGGT TGCTC AGGGT GAGGC
100
GGAAA AGAGT CGAOC CAACC TTCOC CGGGC CGTCA CTGGC ACAGG CTGGA
150
CAGCA AAAGG GCAGA TCACA GTGCT GGACA TGCAC CCAGG CTCTG GGAAG
200
ACCCA CAGAG TOCTC CCGGA GGTCA TTGGC CAATG CATTG ACAGA CGCCT
250
AAGGA CATTG GTGTT GGGCC CAACC CGTGT GGTGC TTAAG GAAAT GGAGC
300
GTGCC TTGAA TGGGA AGAGG GTCAG GTTCC ATTCT COTGC AGGTC GACGG AT

```

Fig. 1 Nucleotide sequence of the single-stranded DNA fragment (feDNA) (ref. 12).

Methods. The feDNA cloned in M13mp7 bacteriophage as described previously¹² was isolated and 3'-end labelled by the Klenow fragment of *E. coli* DNA polymerase I (PolIk) with [α -³²P]-dNTP by the method in ref. 13. This method involves *Bam*HI cutting the single-stranded phage DNA, forming a double-stranded region with the *Bam*HI site close to the point of insertion of a foreign DNA fragment. The ³²P-labelled DNA fragment electrophoretically purified was electroeluted from the gel, desalted and precipitated following a standard procedure¹⁴. The conversion of the feDNA to the partially double-stranded DNA (in the region 1–259) was carried out with DNA polymerase using TGACCCTCTCTCCCAT as a primer complementary to the feDNA in positions 259–274. Reaction mixture (100 μ l) containing 50 μ M of four non-radioactive dNTPs, 10 nM ³²P-labelled feDNA and 2 μ M primer in 50 mM NaCl, 10 mM MgCl₂, 1 mM dithiothreitol and 10 mM Tris-HCl, pH 7.5, was incubated in the presence of 0.5 unit of PolIk for 10 min at 25 °C. The reaction was stopped by addition of 10 μ l of 150 mM EDTA, pH 8.3, and the products were resolved by gel electrophoresis under non-denaturing conditions. The duplex, which has a higher electrophoretic mobility than the feDNA, was isolated from the gel. *Pst*I (nucleotides 12–17) and *Msp*I (nucleotides 166–169) restriction endonucleases hydrolysed the product demonstrating that it is the expected dsDNA fragment.

However, DNA degradation by radicals formed in the metal-catalysed auto-oxidation of ascorbate (or other reducing agents) is well known^{8–11}. In the past, the observed cleavage reactions have not been very specific. Here, we report a non-enzymatic cleavage of single-stranded DNA occurring at unique sites due to redox reactions involving copper. This could be considered a 'self-cleavage' reaction, by analogy with the lead-induced non-redox RNA cleavage reaction^{3–5}. This site-specific cleavage of DNA, stimulated by ascorbate and hydrogen peroxide, is most efficient under physiological conditions, so this phenomenon may have biological significance.

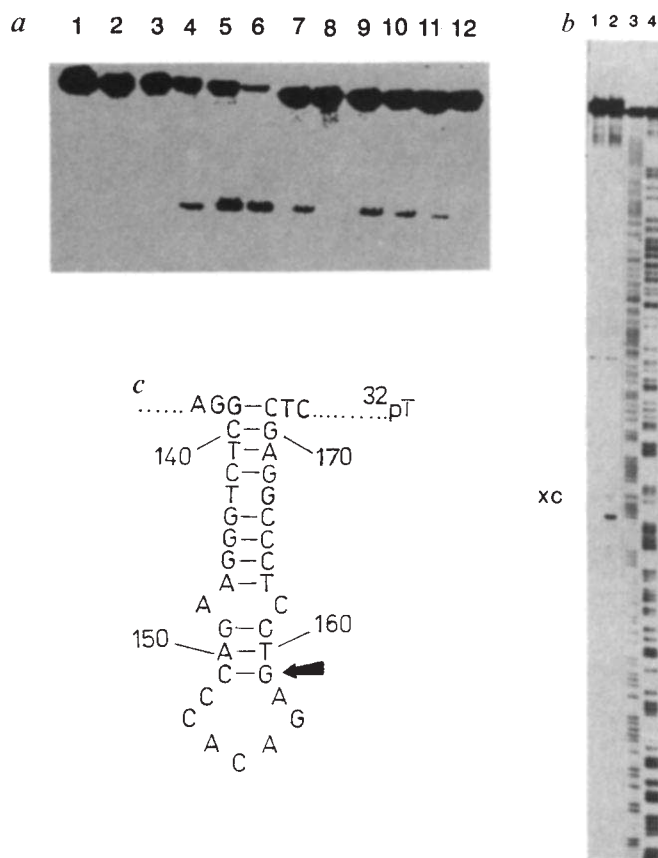
Our observations were made in experiments on the oxidative damage of a single-stranded DNA fragment (302 nucleotides, named feDNA) corresponding to a part of the tick-borne encephalitis viral RNA sequence (Fig. 1). Electrophoretic analysis showed that feDNA incubated with 'ascorbate plus H₂O₂' was cleaved mainly at a single site (Fig. 2). When poly(I)—or poly(A)—used as a free-radical trap was omitted from the buffer, some minor random degradation of the feDNA occurred, as well as the site-specific reaction. Removal of trace amounts of the free metal ions, which participate in redox reactions, from the reaction mixture with 1–2 μ M EDTA and the use of the fresh double-distilled water greatly reduced this non-specific cleavage. Comparison with Maxam–Gilbert sequencing ladders run on the same gel enabled us to identify the exact site of cleavage at a G nucleotide at position 159 (G159) (Fig. 2b). No substantial cleavage was observed when feDNA was incubated with either ascorbate or H₂O₂ alone (Fig. 2a).

Micromolar concentrations of Cu²⁺, in the reaction mixture of feDNA with 'ascorbate plus H₂O₂', accelerated the site-specific cleavage of this DNA (Fig. 2a). Copper ions also exhibited this specific activity in the presence of ascorbate alone, but the combination of H₂O₂ and copper ions was inactive (Fig. 2a).

To determine whether other metal ions are cofactors in the ascorbate-mediated site-specific cleavage of feDNA, various biologically important and toxic transition metal ions, such as

Fig. 2 Site-specific cleavage of the single-stranded feDNA. *a*, Autoradiogram of 10% polyacrylamide gel electrophoretic analysis of 32 P-labelled 10 nM feDNA incubated without reagents (lane 1) and in the presence of 1 mM ascorbate (lane 2); 1 mM H_2O_2 (lane 3); 1 mM ascorbate and 1 mM H_2O_2 (lane 4, 9–12); 1 mM ascorbate, 1 mM H_2O_2 and $1 \mu\text{M}$ Cu^{2+} (lane 5) or $2 \mu\text{M}$ Cu^{2+} (lane 6); 4 mM ascorbate and $4 \mu\text{M}$ Cu^{2+} (lane 7); 4 mM H_2O_2 and $4 \mu\text{M}$ Cu^{2+} (lane 8). Some samples of feDNA were preincubated for 2 min at 95°C in water (lane 9) or in EDTA (pH 8.0) at the following concentrations: $1 \mu\text{M}$ (lane 10), $10 \mu\text{M}$ (lane 11) and $100 \mu\text{M}$ (lane 12). XC, position of xylene cyanol dye. *b*, Autoradiogram of 5% polyacrylamide gel electrophoretic analysis of 32 P-labelled 10 nM feDNA incubated without reagents (lane 1) and in the presence of 1 mM ascorbate plus 1 mM H_2O_2 (lane 2). Lanes 3 and 4 are 'A+G'- and 'C+T'-specific cleavages¹⁴ of feDNA, respectively. *c*, The cleavage site (indicated by the arrow) in a potential hairpin structure of feDNA between nucleotide positions 130 and 170.

Methods. Redox cleavages were carried out for 60 min in buffer LCH (0.15 M LiCl, 0.05 M HEPES, pH 7.5, and 1 mg ml^{-1} poly(1)) at 25°C . Reactions were terminated by adding 20 volumes of 2% LiClO_4 solution in acetone. The precipitate was collected by centrifugation, twice washed with ethanol, vacuum dried and dissolved in 80% formamide, 1 mM EDTA and 50 mM Tris-borate, pH 8.3, with marker dyes. All the chemicals were of the highest grades available. Every reagent solution was prepared immediately before use in fresh double-distilled water. The potential secondary structure of the feDNA was predicted by computer analysis and verified by the cleavage of the single-stranded DNA with S1 nuclease and alkylating oligonucleotide derivatives as described in ref. 12.



Be^{2+} , Mg^{2+} , Al^{3+} , Ti^{3+} , Cr^{3+} , Mn^{2+} , Fe^{2+} , Co^{2+} , Ni^{2+} , Zn^{2+} , Ga^{3+} , Pd^{2+} , Ag^+ , Cd^{3+} , In^{3+} , Sn^{2+} , Pt^{2+} , Au^{3+} , Hg^{2+} , Ti^+ , Pb^{2+} , Bi^{3+} , Ce^{3+} , Pr^{3+} and Eu^{3+} , were examined under comparable conditions. None of these ions were as specifically active as Cu^{2+} . In competition experiments, these metal ions (except for Au^{3+}) at 5–25 μM did not decrease the rate of the site-specific cleavage of feDNA in buffer LCH containing 5 μM Cu^{2+} and 'ascorbate plus H_2O_2 ' (both at 1 mM) regardless of whether they were added before or after Cu^{2+} .

The preincubation of feDNA with such a strong chelator as EDTA completely abolished the specific cleavage activity of 'ascorbate plus H_2O_2 ' (Fig. 2*a*) only at an EDTA concentration higher than that of trace Cu^{2+} (always considered a contaminant of aqueous DNA solutions^{10,17}) by a factor of 10^2 – 10^3 . As the Cu^{2+} –EDTA complex is not capable of inducing the DNA scission⁹, the inhibitory effect seems to involve the competition of EDTA with feDNA for copper ions. Thus, the site-specific cleavage of feDNA requires copper ions to be bound to the DNA in a specific way. In the absence of strong chelators, feDNA can capture copper ions from solution.

Very interesting observations concerning the nuclease-like activity of 1,10-phenanthroline–copper ion complexes have been reported^{15–20}. 1,10-Phenanthroline (OP) is known to activate copper-dependent redox reactions by facilitating the reduction of Cu^{2+} to Cu^+ due to the reversible formation of a complex with a copper ion²¹. Therefore, the effect of OP on the cleavage of feDNA by 'ascorbate plus H_2O_2 ' was investigated. It was found that 1 μM OP enhances dramatically the site-specific cleavage rate of feDNA (Fig. 3). However, at high concentrations, OP inhibited site-specific cleavage of feDNA and caused cleavage at some new sites (Fig. 3). The enhancement of site-specific cleavage of feDNA at lower OP concentrations may be attributed to the formation of a mixed complex feDNA–Cu–OP. A similar ternary complex, RNA–Fe–OP, has been observed previously²². The inhibition of the site-specific cleavage of feDNA at higher OP concentration is obviously due to the extraction of the copper ion from the metal-binding centre as

a $\text{Cu}(\text{OP})_2$ complex. As $\text{Cu}(\text{OP})_2$ also has a nuclease-like activity in the presence of ascorbate and H_2O_2 , this complex induces a multiple feDNA scission, showing some site-specificity. At OP concentrations above 1 mM, the latter process also disappears (Fig. 3).

Thus, our experiments demonstrate that copper ions, tightly and specifically bound to feDNA molecules, are available for interaction with ascorbate, H_2O_2 or OP. As a consequence, highly reactive 'crypto-OH-radicals' may be generated by known redox reactions^{23–26}. The highly specific nature of the cleavage reaction of feDNA confirms that hydroxyl radicals formed at the point of copper binding are not freely diffusible and that the reaction proceeds at the moment of formation. In accordance with the known mechanism of DNA damage by OH-radicals¹⁶, feDNA was not only directly cleaved in this specific reaction, but alkali-labile lesions were observed as well (after 1 M piperidine or 0.1 M LiOH treatment at 90°C) at the same site (data not shown).

Certainly, numerous 'copper-ion-binding sites' are present on each feDNA molecule, but only one, 'the active centre', meets the rigorous stereochemical requirements which greatly facilitate the specific cleavage reaction. This site can be considered as a local metallo-enzyme. The rate of the site-specific cleavage was measured at different temperatures (Fig. 4*a*). A sharp decrease of this rate in the temperature range of 50 – 60°C suggests that a feature of secondary or tertiary structure of the feDNA, which is denatured at high temperatures, is essential for the site-specific cleavage. Figure 2*c* shows a section of a possible feDNA secondary structure containing the scission point. It is interesting to note that the cleavage occurs in the only stable stem-and-loop structure predicted in the feDNA molecule. The essential role of the feDNA structure is also suggested by experiments where the feDNA was converted to a double-stranded DNA fragment (see Fig. 1). The treatment of the duplex DNA molecule by 'ascorbate plus H_2O_2 ' with or without OP under the same conditions as the single-stranded DNA did not lead to site-specific cleavage of the fragment. Therefore, the highly specific cleavage

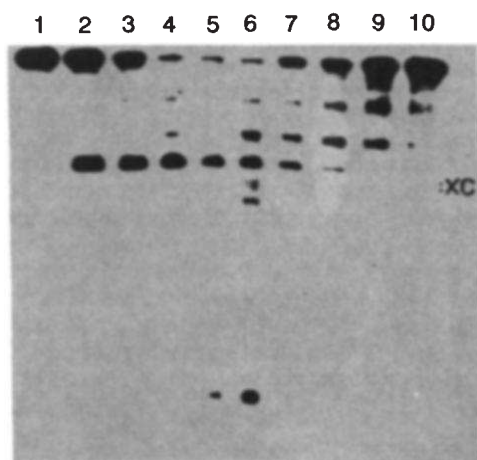


Fig. 3 Autoradiogram of 10% polyacrylamide gel electrophoretic analysis of ^{32}P -labelled 10 nM feDNA incubated without OP (lane 1) and in the presence of OP at the following micromolar concentrations: 1, 5, 10, 20, 50, 100, 250, 500 and 1000 μM (lanes 2–10, respectively).

Methods. Reactions were carried out in the presence of 1 mM ascorbate and 1 mM H_2O_2 for 1 min in buffer LCH at 25 $^\circ\text{C}$.

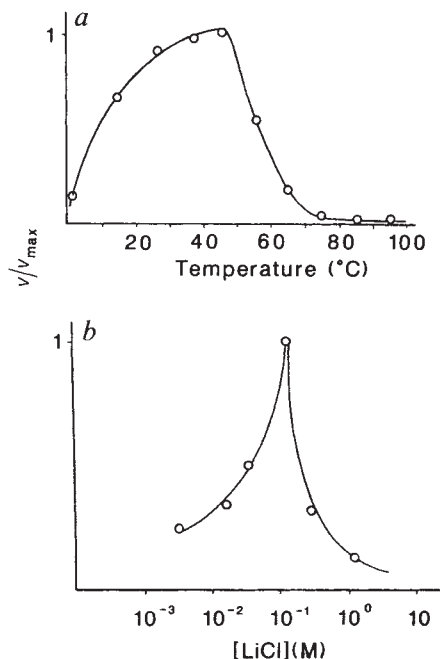


Fig. 4 Effects of *a*, temperature and *b*, ionic strength on the relative rate of the site-specific cleavage of feDNA (at 10 nM concentration) in the presence of 'ascorbate plus H_2O_2 ' (both at 1 mM) and 2 μM Cu^{2+} .

Methods. Reactions were carried out for 15 min, *a*, in buffer LCH, *b*, in 5 mM HEPES (pH 7.5). The cleavage products were resolved by electrophoresis through 10% polyacrylamide gel. After autoradiography, the radioactive bands were cut of the gel and counted by liquid scintillation.

observed is a property of the structure of the single-stranded form.

We revealed a sharp increase in the rate of site-specific cleavage of feDNA under physiological pH values (data not shown) and ionic strengths (Fig. 4*b*). These observations and the fact that ascorbate, H_2O_2 and copper ions participate in the metabolism of all aerobic cells^{27,28} suggest that we have identified a phenomenon of biological significance.

We found a similar single-site redox cleavage of single-

stranded DNA corresponding to a part of Rous sarcoma viral RNA sequence (data not shown). As the choice of both DNAs examined was rather arbitrary, the phenomenon of copper-dependent redox cleavage of a single-stranded DNA at a unique site might be widespread. A similar process may occur in RNAs and double-stranded DNAs which can form local stem-and-loop (or palindromic) structures. It is known that DNA isolated from various sources contains 'endogenous' copper ions and it has been speculated that these ions are bound specifically to disturbed regions of the DNA structure²⁹.

Active copper-binding sites in DNA structures may be related to the phenomenon of 'hot' mutation points. The copper-dependent specific redox damage of DNA may be one of the mechanisms explaining antiviral³⁰, antitumor³¹ and toxic^{27,28} effects of ascorbic acid which are known to be enhanced in the presence of copper ions. This DNA damage activity may also contribute to the known toxic and mutagenic action of copper ions²⁹.

Received 9 February; accepted 30 June 1988.

- Werner, C., Krebs, B., Keith, G. & Dirheimer, G. *Biochim. biophys. Acta* **432**, 161–175 (1976).
- Rordorf, B. F. & Kearns, D. R. *Biopolymers* **15**, 1491–1504 (1976).
- Rubin, J. R. & Sundaralingam, M. *J. Biomol. Struct. Dynam.* **1**, 639–646 (1983).
- Brown, R. S., Hingerty, B. E., Dewan, J. C. & Klug, A. *Nature* **303**, 543–546 (1983).
- Brown, R. S., Dewan, J. C. & Klug, A. *Biochemistry* **24**, 4785–4801 (1985).
- Guerrier-Takada, C., Haydock, K., Allen, L. & Altman, S. *Biochemistry* **25**, 1509–1515 (1986).
- Basile, L. A., Raphael, A. L. & Barton, J. K. *J. Am. chem. Soc.* **109**, 7550–7551 (1987).
- Chiou, S.-H. *J. Biochem.* **94**, 1259–1267 (1983).
- Chiou, S.-H. *J. Biochem.* **96**, 1307–1310 (1984).
- Chiou, S.-H., Chang, W.-C., Jou, Y.-S., Chung, H.-M. M. & Lo, T.-B. *J. Biochem.* **98**, 1723–1726 (1985).
- Hertzberg, R. P. & Dervan, P. B. *Biochemistry* **23**, 3934–3945 (1984).
- Vlassov, V. V., Zarytova, V. F., Kutavi, I. V., Mamaev, S. V. & Podyminogin, M. A. *Nucleic Acids Res.* **14**, 4065–4076 (1986).
- Patton, J. R. & Chae, C.-B. *Analyt. Biochem.* **126**, 231–234 (1982).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
- Marshall, L. E., Graham, D. R., Reich, K. A. & Sigman, D. S. *Biochemistry* **20**, 244–250 (1981).
- Sigman, D. S., Spassky, A., Rimsky, S. & Buc, H. *Biopolymers* **24**, 183–197 (1985).
- Sigman, D. S. *Acc. Chem. Res.* **19**, 180–186 (1986).
- Chen, C.-H. B. & Sigman, D. S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7147–7151 (1986).
- Goldstein, S. & Czapski, G. *J. Am. chem. Soc.* **108**, 2244–2250 (1986).
- Drew, H. R. *J. molec. Biol.* **176**, 535–557 (1984).
- James, B. R. & Williams, R. J. P. *J. chem. Soc.* 2007–2019 (1961).
- Wacker, W. E. C. & Vallee, B. L. *J. biol. Chem.* **234**, 3257–3261 (1961).
- Samuni, A., Chevion, M. & Czapski, G. *J. biol. Chem.* **256**, 12632–12635 (1981).
- Rowley, D. A. & Halliwell, B. *Archs Biochem. Biophys.* **225**, 279–284 (1983).
- Marx, G. & Chevion, M. *Biochem. J.* **236**, 397–400 (1985).
- Morita, J., Ueda, K., Nanjo, S. & Komano, T. *Nucleic Acids Res.* **13**, 449–458 (1985).
- Halliwell, B. & Gutteridge, J. M. C. *Biochem. J.* **219**, 1–14 (1984).
- Weitberg, A. B. *Mutat. Res.* **191**, 53–56 (1987).
- Sisoeff, I., Grisvard, J. & Guille, E. *Prog. Biophys. molec. Biol.* **31**, 165–199 (1976).
- Murata, A., Suenaga, H., Hideshima, S., Tanaka, Y. & Kato, F. *Agric. Biol. Chem.* **50**, 1481–1487 (1986).
- Bram, S. *et al. Nature* **284**, 629–631 (1980).

Elbow motion in the immunoglobulins involves a molecular ball-and-socket joint

Arthur M. Lesk*† & Cyrus Chothia*‡

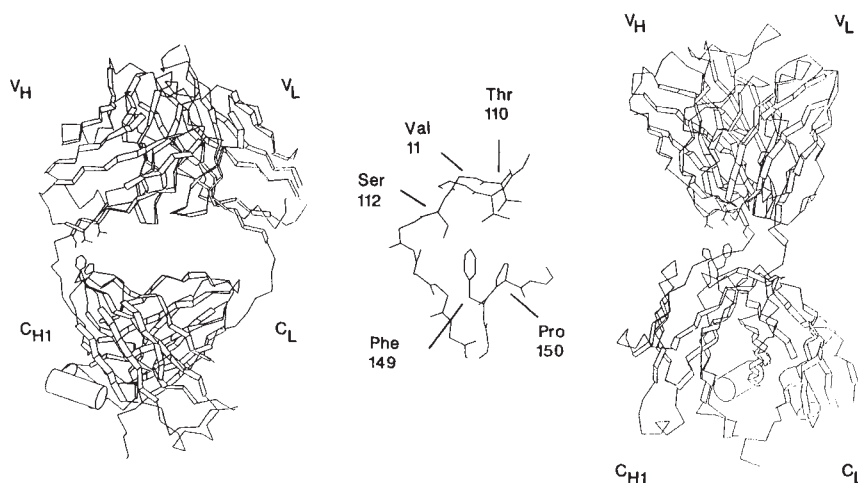
* MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

† European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 1022.09, D-6900 Heidelberg, FRG

‡ Christopher Ingold Laboratory, University College London, 20 Gordon Street, London WC1H 0AJ, UK

Studies by electron microscopy, fluorescence polarization, hydrodynamics and X-ray crystallography have demonstrated the ability of different parts of immunoglobulin molecules to move relative to each other. This movement facilitates the multiple interactions that antibodies make with polyvalent antigens and effector proteins. Comparisons of the atomic structures of immunoglobulins of the same sequence in different crystal environments, and of those with different sequences, have shown that the movements involve local changes in the conformation of the peptides linking different domains. These changes occur in (1) the hinge regions that link the Fab fragment to the Fc, and (2) the switch regions that link

Fig. 1 The conserved V_H to C_{H1} contacts and the switch peptides. The folding of the polypeptide chain in the immunoglobulin fragment Fab Kol⁷ is shown from two approximately orthogonal views. Vertices represent the position of the α -carbon atom of each residue. α -Carbons of the residues that form strands in the β -pleated sheets are joined by ribbons; α -helices are shown as cylinders, and the α -carbons of residues not in secondary structure are joined by a line. Five side chains are shown: three in V_H : 11 Val, 110 Thr and 112 Ser; and two in C_{H1} : 149 Phe and 150 Pro. Contacts between these two sets of residues occur in all the known immunoglobulin structures. The view on the right shows that V_H - C_{H1} contacts are not colinear with the peptides that link the V and C domains. It is these peptides—the switch peptides—that alter their conformation as the orientation the V_L - V_H dimer, relative to the C_L - C_{H1} dimer, changes.



the V_L - V_H dimer to the C_L - C_{H1} dimer¹⁻¹⁵. We show here that in immunoglobulins of known structure, the movement of the V_L - V_H dimer relative to the C_L - C_{H1} dimer also involves the interactions of three V_H and two C_{H1} residues that form the molecular equivalent of a ball-and-socket joint. The almost absolute conservation in the sequences of immunoglobulins and T-cell receptors of the residues that form these interactions suggests that this is a general feature of functional importance.

In this study we used the atomic coordinates produced by the X-ray crystallographic analysis⁵⁻⁹ of the Fab fragments New; McPC603, Kol, J539 and Hy5. The atomic coordinates were obtained from the protein data bank¹⁶ or by personal communication.

The orientation of the V_L - V_H dimer relative to the C_L - C_{H1} is described by the 'elbow angle'. The domain V_L is related to the domain V_H by an approximate twofold axis. The domains C_L and C_{H1} are also related by an approximate twofold axis and the elbow angle is defined as the angle between these two axes¹⁻⁴. In the Fab fragments of known structure the values of the elbow angle⁵⁻¹⁵ vary between about 180° and 130°. Abola *et al.*³ compared the Bence-Jones protein Mcg structures that are found in different crystal forms with different elbow angles. Marquart *et al.*⁷ examined the atomic structures of Fab Kol, where the elbow angle is 170°, and Fab New, where it is 135°. Both studies showed that the differences in the elbow angles involve small differences in the main chain conformation of the three or four residues in the two peptides that link V and C domains (Fig. 1).

We determined which residues make contacts between V and C domains in the five Fab structures. For the Fab fragments New, McPC603 and Kol, we obtained results similar to those given in the original descriptions of their structures⁵⁻⁷, although somewhat different criteria were used for contact residues. We find that the V_H and C_{H1} contacts include a conserved set of interactions involving five homologous residues, plus a few additional contacts that vary from structure to structure. The contacts between the V_L and C_L differ in Fab fragments New, McPC603, J539 and Hy5, and are absent in Fab Kol.

In all five Fab structures, residues 11, 110 and 112 in V_H pack against residues 149 and 150 in C_{H1} (residue numbering according to ref. 17). In the five proteins the residues at these positions are the same or very similar: Leu or Val is found at position 11, Thr at position 110, Ser at position 112, Phe at position 149 and Pro at 150. Residues 11, 110 and 112 are part of the framework β -sheet and their side chains are adjacent in space. Residues 149 and 150 are in a turn of the polypeptide chain (Fig. 1). In all five structures 150 Pro has a *cis* conformation⁵⁻⁷, with its pyrrolidine ring packed against the phenyl ring of 149 Phe. Inspection by computer graphics of atomic models of the Fab structures shows that the three β -sheet V_H residues pack around the two adjacent C_{H1} residues (Fig. 1).

Although the V_H to C_{H1} contacts made by these residues are very similar in the five structures, they are not identical. This is because of differences in the positions of the V_H residues relative to the C_{H1} residues. The extent of the shift between any pair of Fab structures can be found by superposing the C_{H1} residues and calculating the differences in the mean position of the V_H residues. The results of such calculations are shown in Fig. 2. The largest difference is between Fab Kol and Fab McPC603, where the three V_H residues differ in position by a translation of 4.4 Å and a rotation of 36°. These differences in position are related to the differences in the values of the elbow angle: as the elbow angle changes, the V_H residues move over the C_{H1} residues (Fig. 2).

Inspection of space-filling models shows that the relative movement of the residues resembles that which occurs in a ball-and-socket joint, with the two adjacent C_{H1} residues forming the 'ball' and the three β -sheet residues forming the 'socket'. This is illustrated in Fig. 3, which uses space-filling drawings to show how C_{H1} and V_H residues pack together in Fab Kol and Fab McPC603.

An examination of the fifteen values now known for the elbow angle^{3,5-14} shows that they evenly cover a range of some 50°: between 130° and 180°. Values greater than 180° are not found. The residues that form the ball-and-socket joint are not colinear with the switch peptides but displaced to one side (Fig. 1) so that the extension of the elbow angle to values appreciably

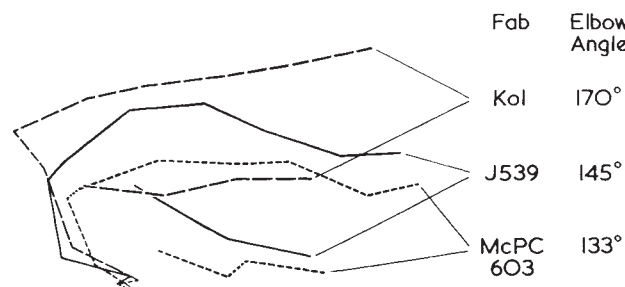


Fig. 2 The correlation between the elbow angle and the position of the V_H residues in the ball-and-socket joint. Three Fab structures, Kol (ref. 7), J539 (ref. 8) and McPC603 (ref. 6) were superposed by a least-squares fit of C_{H1} residues 149 and 150 (not shown), and the positions of the α -carbon atoms of the V_H peptides 10 to 13 and 108 to 120 (which contain the contact residues 11, 110 and 112) were drawn as shown here. The values of the elbow angles in these three Fabs are 170°, 145° and 133° respectively⁵⁻⁸. The position of the peptides and the value of the elbow angle in Fab New are close to those in Fab McPC603. The position of the peptides and the value of the elbow angle in Fab Hy5 are close to those in Fab Kol.

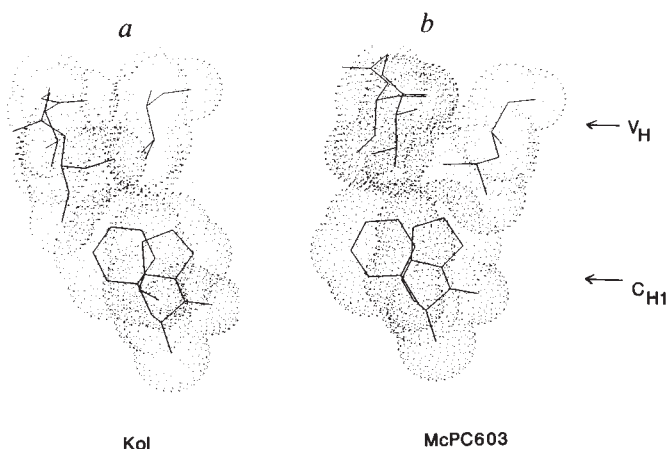


Fig. 3 The ball-and-socket character of the conserved V_H - C_{H1} contact. Surfaces at the van der Waals' radii of the atoms show how the three V_H β -sheet residues pack around the two adjacent C_{H1} residues in *a*, Fab Kol and *b*, Fab McPC603. The elbow angle is 170° in Fab Kol and 133° in Fab McPC603. Comparison of *a* and *b* shows how the change in the elbow angle produces a change in the contact, like that found in a ball-and-socket joint. The positions of the V_H residues in the other three Fab structures are close to, or between, the two positions shown here.

greater than 180° would result in the V_H residues moving away from the C_{H1} . This suggests that the arrangement of the switch peptides and ball-and-socket joint, and the contacts^{5,6} made between V_L - V_H and C_L - C_{H1} at 130° , limit the values of the elbow angle to a range close to that observed in the currently known structures.

The general occurrence and functional importance of the V_H - C_{H1} ball-and-socket joint is indicated by the almost absolute conservation in immunoglobulin sequences of the residues involved. The residues found at positions 11, 110, 112, 149 and 150 in the known V_H and C_{H1} sequences¹⁷ are listed in Table 1.

T cells on their surface have two types of receptors, $\alpha\beta$ and $\gamma\delta$, that bind antigens. Like Fab fragments of the immunoglobulins, they are heterodimers with variable ($V\alpha$, $V\beta$, $V\gamma$ and $V\delta$) and constant ($C\alpha$, $C\beta$, $C\gamma$ and $C\delta$) domains. The comparison of immunoglobulin and T-cell-receptor sequences shows that the corresponding domains of these sets of molecules have structures of similar fold¹⁸⁻²². Examination of the sequences of the T-cell receptors shows that the β -chains in the $\alpha\beta$ receptors and the γ -chains in the $\gamma\delta$ receptors have, at the sites homologous to 11, 149 and 150 in the immunoglobulin heavy chain, the same or very similar residues; and at the sites homologous to 110 and 112 they have the same, or medium-sized, hydrophobic residues (Table 1). The similarity between the residues conserved at these sites and those conserved in the immunoglobulin heavy chains suggests that a flexible ball-and-socket joint also occurs in the structures of the T-cell receptors.

The conserved ball-and-socket joint can be seen as having a dual function in immunoglobulins. It facilitates the motion of V_L - V_H relative to C_L - C_{H1} . It also prevents the formation of a rigid contact between V and C domains that might occur if it were absent.

On a more general level, the ball-and-socket joint provides a flexibility greater than that of packed secondary structures where the side chains interdigitate. The largest observed change in the elbow angle involves a shift of $4.4 \text{ \AA}$ by the main-chain regions of the V_H residues relative to those in C_{H1} (see above and Fig. 2). The main chains of close-packed α -helices are observed to shift no more than 1.5 - $2.0 \text{ \AA}$ relative to each other, and this distance is likely to be close to the limit allowed by low-energy conformational adjustments²³⁻²⁵. In the allosteric transition in haemoglobin the α_1 subunit moves relative to the β_2 subunit by

Table 1 Conservation of the residues involved in the ball-and-socket joint

Position	No. of known sequences	Residues found and their frequencies
Immunoglobulins		
V_H 11	617	Leu:556 Val:24 Phe:11
J_H 110	10	Thr:10
112	10	Ser:10
C_{H1} 149	55	Phe:38 Leu:11
150	55	Pro:55
T-cell $\alpha\beta$ receptors		
$V\beta$ 21	54	Val:26 Ile:24 Leu:3
$J\beta$ 115	12	Thr:13 Ser:4 Leu:3 Ile:1
117	21	Leu:14 Val:4 Thr:2 Ile:1
$C\beta$ 37	4	Phe:3 Tyr:1
38	4	Pro:4
T-cell $\gamma\delta$ receptors		
$V\gamma$ 11	14	Val:8 Ile:4
$J\gamma$ 115	3	Val:2 Ile:1
117	3	Thr:3
$C\gamma$ 38	5	Phe:5
39	5	Pro:5

Data from the sequences collected in Kabat *et al.*¹⁷ and from our unpublished compilation of the known T-cell-receptor sequences.

a larger distance ($6.0 \text{ \AA}$), but a change in packing is involved and intermediate positions are prevented by steric hindrance^{26,27}.

Elbow motion with conformational change in connecting peptides, as observed in the immunoglobulins, has been used as a general model for domain movements in proteins²⁸. It will be interesting to see whether, in those cases where such a general model is applicable²⁴, there are structures like the molecular ball-and-socket joint described here.

We thank Drs S. Sheriff and D. Davies for the coordinates of Fab Hy5, J. Cresswell for the line drawings, our colleagues for comments on the manuscript, and the Royal Society for support.

Received 16 May; accepted 8 August 1988.

- Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3440-3444 (1974).
- Colman, P. M., Deisenhofer, J., Huber, R. & Palm, W. *J. molec. Biol.* **100**, 257-282 (1976).
- Silverton, E. W., Navia, M. A. & Davis, D. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5140-5149 (1977).
- Abola, E. E., Ely, K. R. & Edmundson, A. B. *Biochemistry* **19**, 432-439 (1980).
- Saul, F. A., Amzel, L. M. & Poljak, R. J. *J. biol. Chem.* **253**, 585-597 (1978).
- Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. *J. molec. Biol.* **190**, 593-604 (1986).
- Marquart, M., Deisenhofer, J. & Huber, R. *J. molec. Biol.* **141**, 369-391 (1980).
- Suh, S. E. *et al. Proteins* **1**, 74-80 (1986).
- Sheriff, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 8075-8079 (1987).
- Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. *Science* **233**, 747-753 (1986).
- Colman, P. M. *et al. Nature* **326**, 358-362 (1987).
- Cygler, M., Boodhoo, A., Lee, J. S. & Anderson, W. F. *J. biol. Chem.* **262**, 643-648 (1987).
- Prasad, L., Vandonselaar, M., Lee, J. S. & Delbaere, L. T. J. *J. biol. Chem.* **263**, 2571-2574 (1987).
- Vitali, J., Young, W. W., Schatz, V. B., Sobottka, S. E. & Kretsinger, R. H. *J. molec. Biol.* **198**, 351-355 (1987).
- Davies, D. R. & Metzger, H. A. *Rev. Immun.* **1**, 87-117 (1983).
- Bernstein, F. C. *et al. J. molec. Biol.* **112**, 535-542 (1977).
- Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. *Sequences of Proteins of Immunological Interest* 4th edn (Public Health Service, NIH, Washington DC, 1987).
- Chien, Y. *et al. Nature* **312**, 31-35 (1984).
- Hedrick, S. M., Nielsen, E. A., Kavalier, J., Cohen, D. I. & Davies, M. M. *Nature* **308**, 153-158 (1984).
- Saito, H. *et al. Nature* **312**, 36-40 (1984).
- Yanagi, Y. *et al. Nature* **308**, 145-149 (1984).
- Saito, H. *et al. Nature* **309**, 757-762 (1984).
- Chothia, C., Lesk, A. M., Dodson, G. G. & Hodgkin, D. C. *Nature* **302**, 500-505 (1983).
- Lesk, A. M. & Chothia, C. *J. molec. Biol.* **174**, 175-191 (1984).
- Elber, R. & Karplus, M. *Science* **235**, 318-321 (1987).
- Perutz, M. F. *Nature* **228**, 726-739 (1970).
- Baldwin, J. & Chothia, C. *J. molec. Biol.* **129**, 175-200 (1979).
- Bennett, W. S. & Huber, R. *Crit. Revs Biochem.* **15**, 291-384 (1984).